

# A crash course on data analysis in asteroseismology<sup>†</sup>

By T.Appourchaux

Institut d'Astrophysique Spatiale  
UMR8617, Université Paris-Sud  
Bâtiment 121, 91405 Orsay Cedex, France

In this course, I try to provide a few basics required for performing data analysis in asteroseismology. First, I address how one can properly treat times series: the sampling, the filtering effect, the use of Fourier transform, the associated statistics. Second, I address how one can apply statistics for decision making and for parameter estimation either in a frequentist or a Bayesian framework. Last, I review how these basic principle have been applied (or *not*) in asteroseismology.

---

Throughout human history, as our species has faced the frightening, terrorising fact that we do not know who we are, or where we are going in this ocean of chaos, it has been the authorities – the political, the religious, the educational authorities who attempted to comfort us by giving us order, rules, regulations, informing – *forming* in our minds – their view of reality.

To think for yourself you must question authority and learn how to put yourself in a state of vulnerable open-mindedness, chaotic, confused vulnerability to inform yourself.

Timothy Leary in *Sound Bites from the Counter Culture* (1989)

## 1. Introduction

This paper attempts to provide a summary of the course I gave during the 25<sup>th</sup> Canaries Island Winter School. In no way, this *course* should be perceived as the final answer to a problem. I hope that this *course* can serve as a basis for students, fellow scientists to go beyond what is written here. As in many approaches that I have pursued, this work is a snapshot of where I am and hopefully a possible starting point from which one can expand to other paths not yet ventured.

This *course* starts with a short historical introduction on signal processing and statistics or how our forefathers started doing data analysis more than 200 years ago. The second part is related to the sampling and acquisition of continuous physical signals for subsequent analysis in a digital world. The third part contains with a broad review of statistics from the so-called *frequentist* and *Bayesian* points of view. The last part is related to the applications of the previous concept to data analysis for asteroseismology, which also includes a description of the physics behind that latter terms.

## 2. Historical overview

The basic principle of asteroseismic data analysis can be summarised as follows:

- acquire signal from a finite world
- compute the Fourier transform of the discrete signal

<sup>†</sup> Lecture notes given at the 22<sup>th</sup> Canary Islands Winter School, November 2010.

- extract the characteristics of the harmonic signals

The first step is closely related to approximation of continuous function using decomposition on a base of orthogonal functions. The latter could be the sine and cosine functions used in the second step, that is Fourier decomposition or transform. The last step is related to inference based on statistics or applications of probability to the data. Hereafter, I will try to place in a historical perspective all of these steps. The perspective will be presented in a chronological order for each subject. This historical review does not pretend to be complete as I am not an epistemologist. The main goal of this historical review is to give the reader some keys on reflecting on some tools that we regularly use. The reader will then be free to delve on the subject or leave it aside.

### 2.1. *Spectral analysis and Digital signal processing*

The father of spectral analysis is the well known Joseph Fourier. He pioneered the decomposition of an arbitrary function in cosine and sine functions in his memoir on heat. In Chapter 3 of his memoir (p. 257), Fourier (1822) expressed the decomposition in harmonic functions that later became, using Euler's notation, the complex Fourier transform. In the original formulation of Fourier lies explicitly the potential for a finite summation. In other words, any arbitrary function can be approximated with a finite summation over the Fourier coefficients, which is the Discrete Fourier Transform (DFT). In essence, this is the introduction of a representation of a continuous world using a finite set: a *digital* world. The expression provided by Fourier was already a digital description of the world. The use of sine and cosine functions was very much at the center of the mathematical world at the beginning of the 19<sup>th</sup> century. In 1805, Carl F. Gauss devised an algorithm for interpolation of cosine and sine functions which would later be recognised as the Fast Fourier Transform (FFT), work which was posthumously published (Gauss 1866). This work was published in Latin but translation can be found in Goldstine (1977) and Heideman et al. (1985). The proposal 27 of Gauss' work is what we now call the FFT (Heideman et al. 1985). The *original* algorithm was discovered (not to say re-discovered) by James Cooley and John Tuckey in 1965, while they were working at IBM and the Bell Telephone Laboratory, respectively (Cooley & Tuckey 1965). As anticipated by Gauss, they showed that the DFT could be speeded up by using the fact that the number of points  $N$  in the transform could be expressed as a products of prime numbers, thereby speeding the computation by  $O\left(\frac{N}{\log N}\right)$ .

At the time of the publication of the paper by Cooley & Tuckey (1965), spectral analysis had already entered the *modern* age of the digital world. When working at the AT&T Bell Telephone Laboratory, Claude (Shannon 1949) introduced the so-called sampling theorem which is key for reducing a continuous finite-bandwidth signal to a digital sample. The frequency at which the signal should be sampled was derived by Nyquist (1924), hence bearing the name the Nyquist frequency (Harry Nyquist belonged to what became later the Bell Telephone Laboratory). The sampling theorem is at the basis of all digital audio equipment since the invention by Sony of the digital audio disc in 1972, later to become the Compact Disc of Philips in 1978. Digital signal processing (DSP) is used in many applications such as speech recognition, compression, audio sampling, home cinema and of course in scientific processing.

It is rather surprising to realise that the existence of the digital world dates back to the beginning of the 19th century. In a way, it is not so strange that we need to describe an infinite world with a finite set of data, for instance using function interpolation. Being ourselves *finite* or limited in *time* and *space*, our world was bound to become sooner or later digital.

## 2.2. Probability, statistics and inference

Probability and statistics are related to one another. Probability provides the mathematical foundations for assessing the chance that a random event will occur. While statistics using probability theory provides inference on what has been *really* observed. Probability theory started with Jakob Bernoulli who was applying combinatorial analysis for calculating probabilities related to the games of chance. He published what is known as the Bernoulli distribution and also introduced the law of large numbers (Bernoulli 1713). The same Bernoulli distribution was approximated by De Moivre (1718) which was a special case of the Central Limit theorem. Later on, Reverend Thomas Bayes solved a problem that was left untouched by De Moivre (1718) related to the probability of occurrence of unrelated events. Proposition 5 of Bayes (1763) is what is known today as the *Bayes theorem*. This theorem was also found independently by Laplace (1774) who was working on the same subject. Unfortunately, this view on probability quite advanced at the times of Bayes and Laplace was not used until it was *re-discovered* by Jeffreys (1939).

Inference is related to how one can deduce from data a theoretical model of the world being observed (with error bars on this model). The origin of the first inference can be traced back to the work of Laplace, related to the use of the arithmetic mean (Laplace 1774). Gauss demonstrated, using the *Maximum Likelihood Principle*, that an estimate of a parameter measured many times can indeed be expressed as an arithmetic mean of these observations (Gauss 1809). A typical inference called *Least Squares*<sup>†</sup> *minimisation* was used by Legendre (1805) for deriving the orbit of comets, and for verifying the length of the meter through the measurement of the Earth's circumference. This technique was also found before Legendre by Gauss but was published until later (Gauss 1809). Gauss also derived what was called the law of errors: the so-called Gaussian distribution of errors (Gauss 1809). The use of this ubiquitous error distribution is a simple consequence of the principle of the *Maximum Entropy Distribution* (Jaynes 2009); the distribution simply reflecting the state of our knowledge (or lack of) by knowing the mean value  $\mu$  and the root means square deviation  $\sigma$  of a set of observations. The principles behind *maximisation* are at the very heart of inference. While Gauss introduced the concept for likelihood, Fisher set the proper mathematical background and theory behind the use of *Maximum Likelihood Estimators*, notably their asymptotic properties and their information content (Fisher 1912, 1925). Around that time, a controversy between Ronald Fisher and Harold Jeffreys marked the start of different view on probability and statistics: *frequentists* vs *Bayesian*. Jeffreys (1939) was key in reviving the approach touched upon by Bayes and Laplace, an approach that was not the main stream of statistical thinking at the time of Fisher. At that point in time, the way was open for Bayesian probability and statistics to be applied in various fields of physics and astrophysics. Jaynes (2009) provides many possible applications of Bayesian approaches such as one used for Fourier analysis.

Now in the 21<sup>th</sup> century, it would be naive to believe that inference can only be based on Bayesian approaches, it is surely not a *panacea*. It should be borne in mind that as much as we evolve, any field of Science evolves accordingly. Even in statistics the evolution is not finished. The current stream is to try to reconcile the various approaches promoted by Fisher and Jeffreys (Berger 1997). Therefore, it is the responsibility of any physicist or astrophysicist to follow this evolution. It is perhaps superfluous to remind the reader that all inferences on the world as we see it come from instrumentation, observation, phenomena that are neither perfect nor deterministic. Let me remind you of a quote by Henri Poincaré: *From this point of view all the sciences would only be unconscious*

<sup>†</sup> translated literally from *Moindres quarrés*

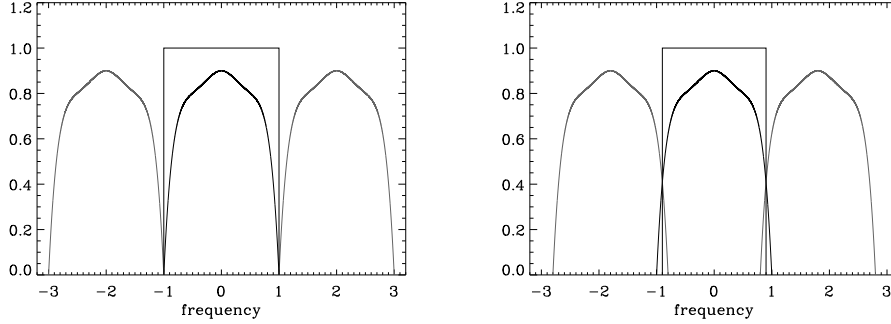


FIGURE 1. The frequency response of a finite bandwidth signal with the boxcar on top, properly sampled according to Shannon theorem (Left), undersampled and aliased (Right)

*applications of the calculus of probabilities. And if this calculus be condemned, then the whole of the sciences must also be condemned (Poincaré 1914).*

### 3. Digital signal processing and spectral analysis

#### 3.1. Time series sampling

Shannon (1949) provided the theorem which allows to sample a continuous signal whose frequencies are contained in a finite bandwidth. Using the decomposition in Fourier series, Shannon (1949) wrote that: *If a function  $x(t)$  contains no frequencies higher than  $\Delta\nu$ , it is completely determined by giving its ordinates at a series of points spaced  $1/(2\Delta\nu)$  apart*", then he wrote:

$$x(t_n) = \int_{-\Delta\nu}^{+\Delta\nu} X(\nu) e^{i2\pi\nu t_n} d\nu \quad (3.1)$$

where  $t_n = n\Delta t$  (with  $\Delta t = 1/2\Delta\nu$ ),  $x(t)$  is the function to be sampled,  $X(\nu)$  is the Fourier transform of  $x(t)$ . From this theorem, and the use of Fourier decomposition, one can then write:

$$X(\nu) = \Pi(2\Delta\nu) \left[ \Delta t \sum_{n=-\infty}^{n=+\infty} x(t_n) e^{i2\pi\nu t_n} \right] \quad (3.2)$$

where  $\Pi$  is the boxcar. The term between brackets is simply the original Fourier decomposition which is implicitly periodic, hence the use of the boxcar for delimiting the frequency space. This is the case represented by the left hand side of Figure 1. Using the inverse Fourier transform of Eq. (3.2), one can show that we can fully reconstruct  $x(t)$  by writing:

$$x(t) = \sum_{n=-\infty}^{n=+\infty} x(t_n) \text{sinc} \left( \frac{t - t_n}{\Delta t} \right) \quad (3.3)$$

where  $\text{sinc} (= \sin x/x)$  is the sinus cardinal function. This equation shows that one can recover *perfectly* a continuous function using samples of that function at regular spacing, whose cadence is provided by the spectral content of that function. In practice, the recovery can only be approximated as the summation over  $\infty$  is impractical.

#### 3.2. Aliasing

There are two cases that provide unwanted frequency leaking into the original spectrum:

- Undersampling of a band-limited signal

- Non band-limited signal

This is what is called *aliasing* whose effect is shown on the right hand side of Fig. 1. In this case high frequency signal leaks, back or aliases, at frequency  $\nu_{\text{alias}} = \nu_{\text{true}} - \Delta\nu$ . Undersampling occurs when the sampling time is larger than Shannon's sampling time ( $1/2\Delta\nu$ ). Undersampling of a band-limited signal can be easily resolved by applying Shannon's theorem. The case of signal having non-limited frequency content is clearly not covered by Shannon's theorem. In that case, there are two techniques that can provide a reduction of the aliasing power: integration and weights.

I shall focus on the effect of integration that is often encountered either when observing time series or making images with discrete arrays such as Charge Coupled Devices (CCD). When one integrates a signal  $x$  over some time ( $\Delta t$ ) I can write:

$$x_{\text{obs}}(t) = \sum_n \left[ \frac{1}{\Delta t} \int_{t_n}^{t_n + \Delta t} x(t) dt \right] \quad (3.4)$$

where  $x$  is regularly sampled at  $\Delta t_s = t_{n+1} - t_n$ . This equation can be rewritten using convolution as:

$$x_{\text{obs}}(t) = \text{III}(\Delta t_s) * \left[ \frac{1}{\Delta t} (x * \Pi(\Delta t))(t) \right] \quad (3.5)$$

where  $\text{III}$  is the Dirac comb. Since the Fourier transform of a Dirac comb is also a Dirac comb, the Fourier spectrum of the signal can then be written as:

$$X_{\text{obs}}(\nu) = \frac{1}{\Delta t_s} \text{III} \left( \frac{1}{\Delta t_s} \right) [X(\nu) \text{sinc}(\Delta t \nu) e^{i\pi \nu \Delta t}] \quad (3.6)$$

Figure 2 shows the resulting effect of the integration when  $\Delta t = \Delta t_s$ . In practice, when the signal is band limited, integration will introduce a filtering effect of the high frequencies. This effect can only be reduced by having a very short integration time with respect to the sampling or  $\Delta t \ll \Delta t_s$ . The effect of aliasing is also shown on Fig. 2. When the signal is non-band limited, there are two possible solutions for reducing the effect of the high-frequency signal:

- Introduction of a window function in frequency (the  $\Pi$  function)
- Integration at 100% duty cycle ( $\Delta t = \Delta t_s$ )

The first solution requires the combination of the time sample by using the Fourier transform of the  $\Pi$  function which is a sinc in time. This solution can only be partially implemented as it would require a sum over  $\infty$ . This solution is naturally implemented when making images through a telescope using discrete arrays. In that latter case, the highest spatial frequencies are cut off by having the telescope diameter providing a cut-off  $D/\lambda$  half that of the spatial pixel sampling; this is illustrated by the left hand side of Fig. 2. In other terms, there is no aliasing in a telescope when the pixel resolution is half of the telescope resolution, that is two samples per resolution element. De facto, in a telescope, the second solution is also used in combination with the first solution. The second solution, although not perfect, will reduce the amplitude of the high frequency noise in time series.

### 3.3. Filtering

The effect of time series filters can simply be studied by understanding the effect of smoothing on time series. Smoothing can be understood as being the application of a weighting function sliding in time: a convolution. This can be written as:

$$x_{\text{sm}}(t) = (x * w)(t) \quad (3.7)$$

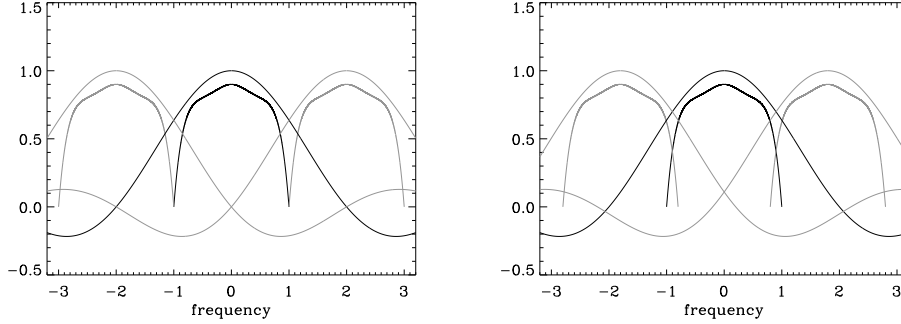


FIGURE 2. The finite bandwidth signal and the sinc function related to the integration over 100% of the sampling time (in black at  $\nu = 0$ ), (in grey at  $\nu = \pm 1/2\Delta\nu$ ); when the signal is sampled according to Shannon's theorem (Left), when the signal is undersampled (Right).

where  $w$  is the weighting function, which can be complex. Using the Fourier transform, this becomes:

$$X_{\text{sm}}(\nu) = X(\nu)W(\nu) \quad (3.8)$$

where  $W$  is the Fourier transform of  $w$ . The smoothing filter typically provides a low pass filter which can be used to derive a high pass filter of the original function  $x$  by computing:

$$x_{\text{fil}}(t) = x(t) - x_{\text{sm}}(t) \quad (3.9)$$

Using the Fourier transform, I get:

$$X_{\text{fil}}(\nu) = X(\nu)(1 - W(\nu)) \quad (3.10)$$

Figure 3 shows the results of applying one time, two times and four times the boxcar on a time series. Applying twice the boxcar is equivalent to a triangular weighting function (convolution of 2 boxcar function), while applying four times the boxcar is equivalent to a bell shape weighting function (convolution of 2 triangle functions). It is rather clear from Fig. 3 that boxcar smoothing should be avoided because it provides too much *ringing* effect of the type Gibbs (1898, 1899) discovered. The introduction of a less sharp transition for all derivatives by multiple boxcar smoothing provides a neat solution to this Gibbs effect.

Another form of filtering is to use the original series shifted in time by  $t_0$  and then subtract the shifted series from the unshifted. In that case the weighting function is simply the Dirac distribution  $w(t) = \delta(t - t_0)$ . Then using Eq. (3.10), the modulus of the filter is then simply

$$|1 - W(\nu)| = 2 \sin(\pi\nu t_0) \quad (3.11)$$

The frequency at which the transmission is half is given by  $\nu_{\text{cuton}} = 1/3t_0$ . This kind of smoothing filter has been used for the data of the Global Oscillation Network Group (GONG) using the first difference obtained with  $t_0 = \Delta t_s$  (Appourchaux et al. 2000).

### 3.4. Time limits and the Discrete Fourier Transform

The Fourier transform is not applicable when doing *real* data analysis. There is no way that we can observe an infinite strings of data. We usually observe during a finite time

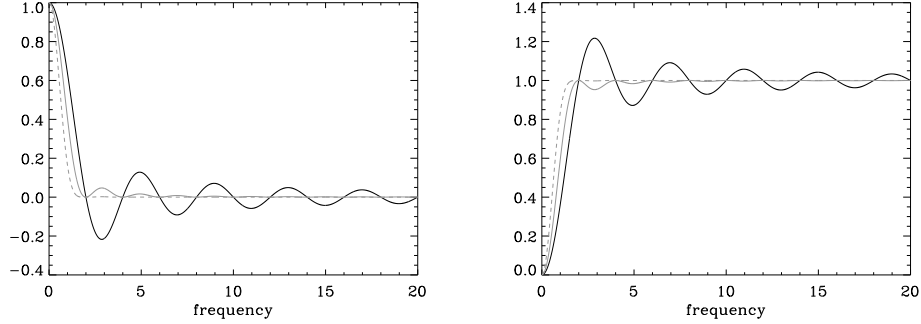


FIGURE 3. (Left) Frequency response of several low pass filter: boxcar (black), triangle or 2 times the boxcar (grey), bell shape of 4 times the boxcar (dashed grey). (Right) Frequency response of several high pass filter resulting from the previous low pass filters corresponding to the same color coding.

$T$  for which we can compute the following Fourier transform:

$$X_T(\nu) = \int_{-T/2}^{+T/2} x(t) e^{i2\pi\nu t} dt \quad (3.12)$$

which can be rewritten as:

$$X_T(\nu) = [X * \text{sinc}(T)](\nu) \quad (3.13)$$

The convolution function of the term in bracket is the sinc function. Figure 4 shows the sinc function for adjacent frequency spaced at  $\pm \frac{1}{T}$ . It is obvious that the spectrum is correlated between the various frequency bins. I will show later that the correlation is in fact null for frequency bins separated by integer values of  $\frac{1}{T}$ , and for slowly varying power spectra. If I combine, the finite observation with a finite bandwidth signal, I then have the Discrete Fourier Transform (DFT):

$$X_{\text{DFT}}(\nu_p) = \Delta t \sum_{n=1}^{n=N} x(t_n) e^{i2\pi\nu_p t_n} \quad (3.14)$$

with  $\nu_p = \frac{p}{T}$ ,  $t_n = n\Delta t_s$ ,  $T = N\Delta t_s$ . Equation (3.14) is simply the truncated original Fourier series, which is then by definition periodic with period of  $\Delta\nu = \frac{1}{t_s}$ . This property directly gives that  $N$  is also the maximum value of  $p$ . The DFT can be computed using the definition given above but it is rather time consuming as the time for computing the summation scales as  $N^2$ . As mentioned in Section 2., Cooley & Tuckey (1965) provided a faster way based on the factorisation properties of the Fourier transform; in that case the time for computing the summations scales as  $N \log N$ .

### 3.5. Fourier transform of stationary processes

The application of Fourier transforms to non deterministic or random functions is key in doing data analysis in astrophysics. For such a process, I define for the random variable  $x$ , the *power spectral density* as:

$$S_T(\nu) = \frac{1}{T} |X_T(\nu)|^2 \quad (3.15)$$

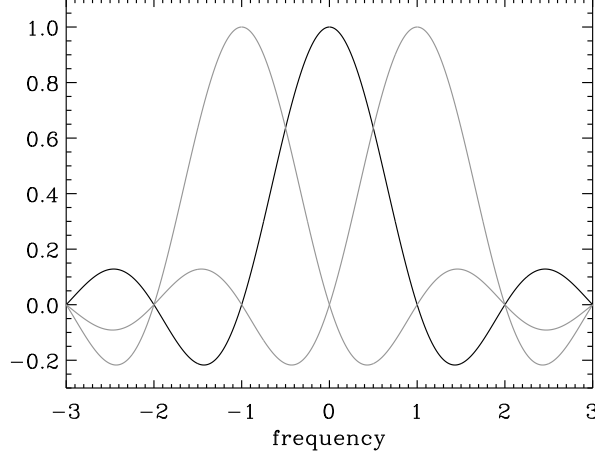


FIGURE 4. The sinc function as a function of frequency normalised to the resolution ( $1/T$ ), for  $\nu = 0$  (black), for  $\nu = \pm 1$  (grey)

where  $X_T(\nu)$  is given by Eq. (3.12). I can then write  $S_T(\nu)$  as:

$$S_T(\nu) = \frac{1}{T} \int_{-T/2}^{+T/2} \int_{-T/2}^{+T/2} x(t)x(t')e^{i2\pi\nu t}e^{-i2\pi\nu t'} dt dt' \quad (3.16)$$

Since the process is stationary I can make the change of variable  $\tau = t - t'$ , and then I have:

$$S_T(\nu) = \int_{-T/2}^{+T/2} \left[ \frac{1}{T} \int_{-T/2}^{+T/2} x(t')x(t' + \tau) dt' \right] e^{i2\pi\nu\tau} d\tau \quad (3.17)$$

The term in brackets is by definition the autocorrelation function ( $C_T(\tau)$ ) of the process taken over a finite window  $T$ . Then I have:

$$S_T(\nu) = \int_{-T/2}^{+T/2} C_T(\tau)e^{i2\pi\nu\tau} d\tau \quad (3.18)$$

Equation (3.18) introduces the so-called Wiener-Khinchin theorem which is essential for understanding the spectral analysis of random processes. As a matter of fact, if I also assume ergodicity of the process (i.e. the temporal average can be exchanged with the spatial average) then I have:

$$C(\tau) = E(x(t)x(t + \tau)) = \lim_{T \rightarrow \infty} C_T(\tau) \quad (3.19)$$

I can then write:

$$S_x(\nu) = \lim_{T \rightarrow \infty} E[S_T(\nu)] = \int_{-\infty}^{+\infty} C(\tau)e^{i2\pi\nu\tau} d\tau \quad (3.20)$$

which is the proper Wiener-Khinchin theorem (Wiener 1930; Khinchin 1934). Using the inverse Fourier transform, I can then also write:

$$C(\tau) = \int_{-\infty}^{+\infty} S_x(\nu)e^{-i2\pi\nu\tau} d\nu \quad (3.21)$$

This relation between the auto-correlation function and the power spectral density is absolutely key in understanding the Fourier analysis of stationary processes.

### 3.6. Parseval's theorem

For  $\tau = 0$ , I also have

$$C(0) = \int_{-\infty}^{+\infty} S_x(\nu) d\nu = E(x^2) \quad (3.22)$$

This equation is also an alternate formulation of the energy conservation principle known as Parseval's theorem (Parseval des Chênes 1806). This important property provided by Eq. (3.22) can be adapted for normalising the DFT. For the DFT given by Eq. (3.14), I can write:

$$\alpha^2 \sum_{p=1}^{p=N} |X_{\text{DFT}}(\nu_p)|^2 = \frac{1}{N} \sum_{n=1}^{n=N} x(t_n)^2 \quad (3.23)$$

where  $\alpha$  is the required normalisation factor used for the spectral density. For Eq. (3.14), it is easy to show that the normalization factor  $\alpha$  is  $1/N$  which is the usual factor used when computing the DFT. If other definitions of the DFT are to be used, the proper normalisation factor  $\alpha$  can be derived with Eq. (3.23). This latter equation is used for calibrating the spectra coming from different routines, all having different normalization factors.

### 3.7. Statistics of the Discrete Fourier Transform

Equation (3.14) is *our bread and butter* for extracting the frequencies associated with periodic phenomenon observed, for instance, in stars. Whenever one carries out observations, one should not forget that the observations come with noise either associated with the observed phenomenon or with the instrument providing the data. Therefore, it is essential to understand the statistics of the Fourier transform of random variables. Let us start with a simple and commonly used example: a random variable  $x$  with an unknown distribution (with  $E(x) = 0$  and  $E(x^2) = \sigma^2$ ) for which all time samples  $x(t_n)$  are independent from each other, in addition the process is assumed to be stationary. Using Eq. (3.14), I can then write the DFT of the  $x(t_n)$  as:

$$X_{\text{DFT}}^r(\nu_p) + iX_{\text{DFT}}^i(\nu_p) = \sum_{n=1}^{n=N} x(t_n) \cos(2\pi\nu_p t_n) + i \sum_{n=1}^{n=N} x(t_n) \sin(2\pi\nu_p t_n) \quad (3.24)$$

where  $X_{\text{DFT}}^r$  and  $X_{\text{DFT}}^i$  are the real and imaginary part of the Fourier spectrum. Since the  $x(t_n)$  are independent and identically distributed (i.i.d.), by virtue of the Central Limit Theorem, the statistical distribution of  $X_{\text{DFT}}^r$  and  $X_{\text{DFT}}^i$  is a normal distribution for  $N \gg 1$  with:

$$E(X_{\text{DFT}}^r(\nu_p)) = E(X_{\text{DFT}}^i(\nu_p)) = 0 \quad (3.25)$$

$$E([X_{\text{DFT}}^r(\nu_p)]^2) = E([X_{\text{DFT}}^i(\nu_p)]^2) = \frac{N}{2}\sigma^2 \quad (3.26)$$

Then, since  $X_{\text{DFT}}^r$  and  $X_{\text{DFT}}^i$  are independent and have the same normal distribution, the statistics of the power spectrum is then by definition a  $\chi^2$  with 2 degrees of freedom (d.o.f.).

Unfortunately (or fortunately) none of the processes that we observe have the properties of being i.i.d. The processes are usually stationary processes but not i.i.d. because these processes have usually a memory such that the correlation of  $x(t_n)$  and  $x(t_m)$  are different from zero when  $t_n \neq t_m$ . Nevertheless, in that case, it can be demonstrated that the components of the Fourier transform are *also* both normally distributed with the same mean of zero and the same variance, which depends upon frequency (Peligrad & Wu

2010). It is amusing to quote Peligrad and Wu on their finding: ‘*In this sense Theorem 2.1 [of Peligrad & Wu (2010)] justifies the folklore in the spectral domain analysis of time series: the Fourier transforms of stationary processes are asymptotically independent Gaussian.*’

### 3.8. Time series sampled at unevenly times

It is quite common in astrophysics to have samples that are not equally spaced in time. This lack of uniformity could be due to a variable number of photon per seconds or due to variable detector read out time. Usually, this can be avoided by carefully designing the electronics, see for instance Fröhlich et al. (1997). Nevertheless, if the time series are unevenly sampled, there are ways and means to find solutions. Equation (3.14) can still be applied, obviously by dropping  $\Delta t_s$ . The problem of the latter equation is that for unevenly times the statistics of the Fourier spectrum is no longer  $\chi^2$  with 2 d.o.f. (Scargle 1982). Equation (3.24) can be adapted such that this statistical property is kept by writing:

$$X_{\text{LS}}^{\text{r}}(\nu_p) = \frac{1}{w(\tau)} \sum_{n=1}^{n=N} x(t_n) \cos(2\pi\nu_p(t_n - \tau)) \quad (3.27)$$

$$X_{\text{LS}}^{\text{i}}(\nu_p) = \frac{1}{v(\tau)} \sum_{n=1}^{n=N} x(t_n) \sin(2\pi\nu_p(t_n - \tau)) \quad (3.28)$$

where  $w$  and  $v$  are given by:

$$w(\tau) = \sum_{n=1}^{n=N} \cos^2(2\pi\nu_p(t_n - \tau)) \quad (3.29)$$

$$v(\tau) = \sum_{n=1}^{n=N} \sin^2(2\pi\nu_p(t_n - \tau)) \quad (3.30)$$

and  $\tau$  is introduced for keeping the invariance in time of the transform given by Eqs. (3.27) and (3.28).  $\tau$  is given by:

$$\tan(2\pi\nu\tau) = \frac{\sum_{n=1}^{n=N} \sin(2\pi\nu t_n)}{\sum_{n=1}^{n=N} \cos(2\pi\nu t_n)} \quad (3.31)$$

The definition provided by Eqs. (3.27) and (3.28) has the benefit of giving a power spectrum or Lomb-Scargle (LS) periodogram  $([X_{\text{LS}}^{\text{r}}(\nu_p)]^2 + [X_{\text{LS}}^{\text{i}}(\nu_p)]^2)$  which is  $\chi^2$  with 2 d.o.f. (Scargle 1982). It must also be pointed out that Eqs. (3.27) and (3.28) are also the solution obtained when applying *Least Squares* minimisation to

$$\sum_{n=1}^{n=N} [x(t_n) - a_c \cos(2\pi\nu t) - a_s \sin(2\pi\nu t)]^2 \quad (3.32)$$

where  $a_c$  and  $a_s$  are given by Eqs. (3.27) and (3.28), respectively. For speed, the LS periodogram is usually computed using the implementation prescribed by Press & Rybicki (1989) which is an approximation of the LS periodogram based upon *extirpolation*<sup>†</sup> on a regular mesh and the use of the FFT. The prescription is then very close to interpolating onto a regular mesh. It is worth noting that most users of the LS periodogram for

<sup>†</sup> Reverse interpolation or extirpolation replaces a function value at any arbitrary point by several function values on a regular mesh

unevenly sampled data are in fact computing the FFT of the original data resampled onto a regular mesh, but with a proper normalization as given by Scargle (1982). It must be noted that the LS periodogram does not provide a better solution to coping with the presence of gaps. The reason is that although the Fourier transform *explicitly* includes gaps as zeros, adding zeros is also *implicitly* performed with the LS periodogram. As a consequence, correlations between frequency bins also exist with the LS periodogram, but these are generally ignored. The correlations in the Fourier transform in the presence of gaps are addressed in the next section.

### 3.9. The influence of gaps in the time series

The impact of the gaps on the Fourier spectrum has been described by Gabriel (1994). The gaps introduce correlation between frequency bins that need to be taken into account when one wants, for example, to fit the power spectrum (See for applications Stahn & Gizon 2008). Hereafter, I will provide the result regarding the correlation of the Fourier spectrum between the frequency bins. Assuming that I observe, a random variable  $x$  through a window  $W$ , the Fourier transform can be written as:

$$\tilde{\mathcal{X}}(\nu) = \int_{-\infty}^{+\infty} x(t)W(t)e^{i2\pi\nu t} dt \quad (3.33)$$

The mean correlation between two frequency bins  $\nu_1$  and  $\nu_2$  is given by:

$$E[\tilde{\mathcal{X}}(\nu_1)\tilde{\mathcal{X}}^*(\nu_2)] = E\left[\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x(t)x(t')W(t)W(t')e^{i2\pi\nu_1 t}e^{-i2\pi\nu_2 t'} dt dt'\right] \quad (3.34)$$

where  $*$  denotes the complex conjugate. Following Gabriel (1993), I have the following properties for the real and imaginary parts of  $\mathcal{X}$ :

$$E[\tilde{\mathcal{X}}_r(\nu_1)\tilde{\mathcal{X}}_r^*(\nu_2)] = E[\tilde{\mathcal{X}}_i(\nu_1)\tilde{\mathcal{X}}_i^*(\nu_2)] \quad (3.35)$$

$$E[\tilde{\mathcal{X}}_r(\nu_1)\tilde{\mathcal{X}}_i^*(\nu_2)] = -E[\tilde{\mathcal{X}}_i(\nu_1)\tilde{\mathcal{X}}_r^*(\nu_2)] \quad (3.36)$$

Using the Fourier transform of  $x(t)$ , I can rewrite Eq. (3.34):

$$E[\tilde{\mathcal{X}}(\nu_1)\tilde{\mathcal{X}}^*(\nu_2)] = \int \int \int \int E[X(\nu)X(\nu')]W(t)W(t')e^{i2\pi[(\nu_1-\nu')t-(\nu_2-\nu')t']} dt dt' d\nu d\nu' \quad (3.37)$$

By construction, I assume that there is no correlation between the real and imaginary parts of the original spectrum  $X$ , that their variances have the same value  $E[X_r^2(\nu)]$  and the frequency bins of the original spectrum are not correlated (See also, Gabriel 1993), such that I have:

$$E[X_r(\nu)X_r(\nu')] = E[X_r^2(\nu)]\delta(\nu' - \nu) \quad (3.38)$$

$$E[X_r(\nu)X_i(\nu')] = E[X_i(\nu)X_i(\nu')] \quad (3.39)$$

where  $\delta$  is the Dirac distribution. Then I can rewrite:

$$E[\tilde{\mathcal{X}}(\nu_1)\tilde{\mathcal{X}}^*(\nu_2)] = 2 \int_{-\infty}^{+\infty} E[X_r^2(\nu)][W(\nu_1 - \nu)W(\nu - \nu_2)]d\nu \quad (3.40)$$

where  $W$  is the Fourier transform of the window function  $w$ . For understanding the impact of Eq. (3.40), let us assume that we observe white noise of mean 0 and of variance  $\sigma_0$  in frequency. In that case, I have:

$$E[\tilde{\mathcal{X}}(\nu_1)\tilde{\mathcal{X}}^*(\nu_2)] = 2\sigma_0^2 \int_{-\infty}^{+\infty} W(\nu_1 - \nu)W(\nu - \nu_2)d\nu \quad (3.41)$$

Using the properties of convolution and the inverse Fourier transform, it can be shown† that I have:

$$\mathbb{E}[\tilde{\mathcal{X}}(\nu_1)\tilde{\mathcal{X}}^*(\nu_2)] = 2\sigma_0^2 \int_{-\infty}^{+\infty} w^2(t)e^{i2\pi(\nu_1-\nu_2)t}d\nu \quad (3.42)$$

The integral in this equation is simply the Fourier transform of the square of the window function, or  $W_{\text{sq}}$ . For a window function such as the one provided by an observing window of length  $T$  (See Eq. 3.12), the correlation is then given by the sinc function. This justifies *a posteriori* the sampling at frequency interval of  $1/T$ , for which the correlation is null. If the variations of  $\mathbb{E}[X_r^2(\nu)]$  are slow with respect to  $W(\nu)$ , I can also rewrite Eq. (3.40) as:

$$\mathbb{E}[\tilde{\mathcal{X}}(\nu_1)\tilde{\mathcal{X}}^*(\nu_2)] \approx 2\mathbb{E}[X_r^2(\nu)]W_{\text{sq}}(\nu_1 - \nu_2) \quad (3.43)$$

where  $W_{\text{sq}}$  is the Fourier transform of  $w^2$ . Of course this approximation does not hold when the variations in frequency are over scales of  $1/T$ . Nevertheless, Eq. (3.43) gives an interesting solution for understanding the correlation between frequency bins in a Fourier spectrum.

It is also useful to understand the correlation between the real and imaginary parts of the Fourier transform. Using Eqs. (3.35) and (3.36), I can derive the very useful formulation:

$$\mathbb{E}[\tilde{\mathcal{X}}_r(\nu_1)\tilde{\mathcal{X}}_r^*(\nu_2)] = \int_{-\infty}^{+\infty} \mathbb{E}[X_r^2(\nu)]\mathcal{R}[W(\nu_1 - \nu)W(\nu - \nu_2)]d\nu \quad (3.44)$$

$$\mathbb{E}[\tilde{\mathcal{X}}_r(\nu_1)\tilde{\mathcal{X}}_i^*(\nu_2)] = \int_{-\infty}^{+\infty} \mathbb{E}[X_r^2(\nu)]\mathcal{I}[W(\nu_1 - \nu)W(\nu - \nu_2)]d\nu \quad (3.45)$$

where  $\mathcal{R}$  and  $\mathcal{I}$  denote the real and imaginary operators, respectively. These two relations can be used when a specific window which is different from the boxcar is used. For example, the use of weights across the observing window will then introduce correlation provided by Eqs.(3.44) and (3.45). The introduction of these weights or tapers is described in the next section.

### 3.10. *Taper estimates of Fourier power spectrum*

Fourier spectrum estimation is well adapted for periodic signals (pure sine waves or stochastic waves) but not necessarily well suited for estimating the spectral density of frequency-dependent noise (pink or red noise). For that purpose, one can:

- average the power spectrum over an ensemble of  $n$  sub-series,
- smooth the power spectra over  $n$  frequency bins or,
- use multitapered spectra using the full time series for deriving a similar average.

Fourier spectrum estimation can be replaced by multitapered spectra that are widely used in geophysics (for a review see Thomson 1982). Multitapered spectra are generated by applying a set of tapers to a single time series, and an estimate of the mean power spectrum is derived from an average of these spectra. Using tapers due to Slepian (1978), the multitapered spectra are statistically independent from one another, and the statistics of the mean spectrum follows a  $\chi^2$  distribution with  $2n$  degrees of freedom (where  $n$  is the number of tapers, Thomson 1982). While the statistics of the average power spectrum (or smoothed power spectrum) also follow a  $\chi^2$  distribution with  $2n$  degrees of freedom: the resolution of the average spectrum is  $n$  times lower than that of the multitapered spectrum. In helioseismology the use of these slepian tapers has been replaced by more practical (but less accurate) sine tapers (Komm et al. 1999). Unfortunately, for

† I leave the demonstration to the reader

sine waves, tapers tend to broaden the peaks, as shown by Thomson (1982). Tapers as such provide more benefit for broader peaks than for narrower peaks.

## 4. Data analysis and statistics

### 4.1. Hypothesis testing

Statistical testing is essential when one wants to decide: *have we found a signal or not?* This is related to *decision theory*, which can be summarised as *how do we choose between one hypothesis versus another in the presence of uncertainties?* In this area, there are two schools of thought: the frequentist school and the Bayesian school.

The difference between a Bayesian and a frequentist relates to their views of *subjective* versus *objective* probabilities. A frequentist thinks that the laws of physics are *deterministic*, while a Bayesian ascribes a belief that the laws of physics are true or *operational*. The *subjective* approach to probability was first coined by De Finetti (1937).

For the rest of us, the difference in views between frequentists and Bayesians can be outlined by taking an example from the Six-nation rugby tournament. For a frequentist, France has been winning over England in their direct confrontation only 39% of these matches since 1906. Based on this result, for a frequentist, France has only 39% chance of winning any future game against England. For a Bayesian, this is a complete different story. A Bayesian may attribute higher chance or lower chance to France to win a given game based on the current physical and technical skills of each player, on the current ability of the players to play as a team and on the psychological and mental health of the players and of the team as a whole. Based on this assessment, a Bayesian might have attributed 70% chance to win the 2010 game, or 20% chance to win the 2011 game. This is an *a posteriori* evaluation of the chances as both games took place while writing this course. It is used as an example.

In short, frequentists assign probability to measurable events that can be measured an infinite number of times, while Bayesians assign probability to events that cannot be measured, like the survival time of the human race (Gott 1994) or like the future outcome of sporting matches.

In what follows, I will try to give an overview of what *I believe* I know on a subject that is rapidly evolving; and what I write is certainly *not* gospel.

#### 4.1.1. Frequentist hypothesis testing

For a frequentist, statistical testing is related to hypothesis testing. In short, we have two types of hypotheses:

- $H_0$  hypothesis or null hypothesis: what has been observed is pure noise
- $H_1$  hypothesis or alternative hypothesis: what has been observed is a signal

For the  $H_0$  hypothesis, I assume known statistics for the random variable  $Y$  observed as  $y$  and assumed to be pure noise; and then set a *false alarm probability* that defines the acceptance or rejection of the hypothesis. The so-called *detection significance* (or *p-value*, terms not widely used in astrophysics) is the probability of having a value as extreme as *the one actually observed*. There is an on-going confusion because statisticians call *the significance level* what astronomers call the *false alarm probability*; and statisticians call the *p-value* what is set in astronomy as the *detection significance* (which is *not* the significance level). Here I shall use the current vocabulary understood in astronomy. For example, the *false alarm probability*  $p$  for the  $H_0$  hypothesis is defined as:

$$p = P_0(\tilde{T}(Y) \geq \tilde{T}(y_c)), \quad (4.1)$$

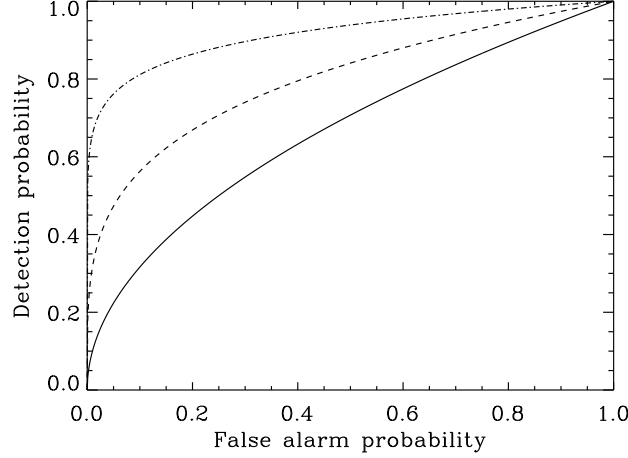


FIGURE 5. Detection probability as a function of the false alarm probability for pure sine waves stochastically excited for various signal-to-noise ratio: 1 (continuous line), 2 (dashed line), 10 (dot-dashed line)

where  $\tilde{T}$  is the statistical test, and  $P_0$  is the probability of having  $\tilde{T}(Y) \geq \tilde{T}(y_c)$  when  $H_0$  is true; and  $y_c$  is the cut-off threshold derived from the test  $\tilde{T}$  and the value  $p$ . For example, take the case of a random variable  $Y$  distributed with  $\chi^2$ , 2 degrees of freedom (d.o.f) statistics, having a mean of  $\sigma$ . If I further assume that  $\tilde{T}(Y) = Y$ , I then have that:

$$p = P_0(Y \geq y_c) = e^{-\frac{y_c}{\sigma}} \quad (4.2)$$

If one observes a value  $\tilde{y}$  of the random variable  $Y$  that is larger than  $y_c$ , the  $H_0$  hypothesis is rejected. The value that is quoted in this case is the *detection significance*  $\mathcal{D}$ , i.e.,

$$\mathcal{D} = e^{-\frac{\tilde{y}}{\sigma}} \quad (4.3)$$

The  $H_0$  hypothesis was used by Scargle (1982) for setting a false alarm probability, and by Appourchaux et al. (2000) to impose an upper limit on g-mode amplitudes. The method was based on the knowledge of the statistical distribution of the power spectrum of full-disc asteroseismic instruments, namely the  $\chi^2$  distribution with 2 d.o.f. For the  $H_1$  hypothesis, I assume given statistics both for the noise and for the signal that we wish to detect, and set a level that defines the acceptance or rejection of that hypothesis.

In this example, I took as given that the test  $\tilde{T}$  was known ( $\tilde{T}(Y) = Y$ ). As a matter of fact, such a test is not obtained in an *ad hoc* manner but can be rationally derived using the Neyman-Pearson lemma. An example of the application of this lemma for deriving the test and level provided by Eq. (4.2) is explained in the next section.

**The Neyman-Pearson lemma.** The derivation of the best test and of the levels associated with the  $H_0$  and  $H_1$  hypotheses is provided by the Neyman-Pearson lemma (Neyman & Pearson 1933). This lemma is very useful for designing tests that will maximise signal detection while minimising noise effects.

LEMMA 1.  $\exists \eta > 0$  such that  $\Lambda(y) = \frac{L(y|H_0)}{L(y|H_1)} \leq \eta$  where  $P(\Lambda(y) \leq \eta | H_0) = \alpha$

where  $L(y|H_0)$  and  $L(y|H_1)$  are the likelihood for each hypothesis and  $\alpha$  is called the *power* of the test. I will show later how one can use such a lemma for a specific case that is often encountered in astronomy: the detection of single frequency peak in a power

TABLE 1. Types of error obtained for different decisions, based upon the statistical test performed, and how the error relates to the status of the  $H_0$  hypothesis.

		Status of $H_0$	
		True	False
Decision	Reject	Type I	Correct
	Accept	Correct	Type II

spectrum of a star having eigenmodes with very long lifetimes. Let us assume that we observe pure noise in a power spectrum, since the statistics is  $\chi^2$  with 2 d.o.f. I can write the likelihood of observing a value  $\tilde{y}$  as:

$$L(\tilde{y}|H_0) = \frac{1}{B} e^{-\tilde{y}/B} \quad (4.4)$$

where  $B$  is the mean noise level in the power spectrum. Next, I assume that the peak is not deterministic but that its amplitude is stochastic with an amplitude  $A$ . The likelihood for  $H_1$  is then:

$$L(\tilde{y}|H_1) = \frac{1}{B+A} e^{-\tilde{y}/(B+A)} \quad (4.5)$$

The likelihood ratio then can be written as:

$$\Lambda(\tilde{y}) = (1+H) e^{-\frac{\tilde{y}H}{1+H}} \quad (4.6)$$

with  $H = A/B$ . Then applying the Neyman-Pearson lemma leads to:

$$\Lambda(\tilde{y}) \leq \eta \Rightarrow \tilde{y} \geq \eta' \quad (4.7)$$

where  $\eta'$  is given by solving

$$P(\tilde{y} \geq \eta'|H_0) = e^{-\frac{\eta'}{B}} = \alpha \quad (4.8)$$

which justifies *a posteriori* the use of Eq. (4.2). Then I can also write the detection probability of a sine waves as:

$$P(\tilde{y} \geq \eta'|H_1) = e^{-\frac{\eta'}{A+B}} = \alpha^{\frac{1}{1+H}} \quad (4.9)$$

Figure 5 shows the result for the detection probability for sine waves stochastically excited. Such a diagram is also called the *receiver operating characteristic* (roc). It provides a very efficient way of assessing the performance of the statistical test used. In summary, the Neyman-Pearson lemma can be used for deriving in a non-arbitrary fashion the best test for accepting/rejecting  $H_0$ . With this lemma, the design of a test is therefore more systematic and less prone to improvisation.

**Is the world dichotomic?** Taking a decision based on the result given by a single test, for either hypothesis, could lead to errors in the decision process. For instance, the null hypothesis could be wrongly rejected when it is true (*false positive* or wrong detection), but could also be wrongly accepted while it is false (*false negative* or no detection in presence of a signal). The *false positive* results in a *Type I* error, while the *false negative*

results in a *Type II* error (see Table 1). The ideal case would be, using the Neyman-Pearson lemma, to set a test that would minimise the occurrence of both types of errors

It has been customary when applying the  $H_0$  hypothesis to set the decision level arbitrarily at 10% (Appourchaux et al. 2000). From the frequentist view point there is nothing wrong in setting *a priori* the decision level before the test is applied. There are three types of result we might obtain from applying the test:

- (a)  $H_0$  always rejected
- (b)  $H_0$  rejected or accepted at a level very close to 10%
- (c)  $H_0$  always accepted

Decision (a) will lead to the mention of a *detection being statistically significant* at a level provided by the *detection significance* (for example from Eq. (4.3)). The question is then to know what was detected. The next step would then be the application of a test for the  $H_1$  hypothesis taking into account assumptions about the detected signal, which may very likely result in the detection of signal. Decision (c) seems straightforward, i.e., noise dominates, but might one then be tempted to lower, *a posteriori*, the decision level? Decision (b) is the more difficult borderline case, forcing us to either accept or reject  $H_0$ . Here, we might ask: are things really that clear cut? What are the chances that if we accept  $H_0$  it is actually wrong (Type II error), or truly right if rejected (Type I error)?

These potential actions result from the application of a frequentist test trying to answer the following question: what is the likelihood of the observed data set  $\tilde{y}$ , given that  $H_0$  is true or  $p(\tilde{y}|H_0)$ ? The *detection significance* mentioned when the test rejects the  $H_0$  hypothesis is nothing but  $p(\tilde{y}|H_0)$ , when actually what we want to know is the likelihood that  $H_0$  is true given the data, i.e.,  $p(H_0|\tilde{y})$  ( $\neq p(\tilde{y}|H_0)$ ). The frequentist view does provide a useful answer when one can repeat the observations ad infinitum. But when we have only one universe, one observation, another approach must be used based upon Bayes' theorem; an approach which in principle gives access directly to  $p(H_0|\tilde{y})$ .

#### 4.1.2. Bayesian hypothesis testing

**On the posterior probability.** We should never forget the *two sides of the coin*: if probability (likelihood) can justify *alone* the rejection or acceptance of an hypothesis, this probability *is not* the significance that the hypothesis is rejected or accepted. The decision levels discussed above are related directly to a well-known controversy in the medical field, concerning improper use of Fisher's p-values as measures of the probability of effectiveness of a medicine or drug (Sellke et al. 2001). The *detection significance* (or p-value) is improperly used as the significance of the evidence against the null hypothesis. It is far from trivial at first sight to understand what is wrong with the *detection significance*. Let us recall the example I gave above for a random variable  $Y$  having a  $\chi^2$  with 2 d.o.f statistics. In that case the *detection significance* is given as:

$$\mathcal{D} = e^{-\frac{\tilde{y}}{\sigma}} \neq P_0(Y \geq \tilde{y}). \quad (4.10)$$

The latter statement ( $\neq$ ) is fundamental. The observation is performed only once providing a value of  $\tilde{y}$  and hence the detection significance  $\mathcal{D}$ . But in no way does it provide the probability that the random variable is *always* above  $\tilde{y}$  (or  $P_0(Y \geq \tilde{y})$ ). It is not correct to assume that if the observation were repeated it would provide the same level  $\tilde{y}$ . The mistake is to ascribe a significance to a measurement performed only once, i.e., not repeated, and spanning just a very small volume of the parameter space (e.g.  $Y \in [\tilde{y}, \tilde{y} + \delta y]$ ). If one makes a measurement  $\tilde{y}$  of the random variable  $Y$  that is above  $y_c$ , the significance of that measurement is *not*  $e^{-\tilde{y}/\sigma}$ . In the framework of Bayesian statistics, we are not interested in the *detection significance* but in the posterior probability of the hypothesis

$p(H_0|\tilde{y})$ , in other words as already stated above  $p(H_0|\tilde{y}) \neq p(\tilde{y}|H_0)$ . A similar description of this misunderstanding has been presented by Sturrock & Scargle (2009).

In order to derive the posterior probability  $p(H_0|y)$ , let us first recall the Bayes' theorem. The theorem of Bayes (1763) relates the probability of an event A given the occurrence of an event B to the probability of the event B given the occurrence of the event A, and the probability of occurrence of the events A and B alone.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (4.11)$$

For example, the probability of having rain given the presence of clouds is related to the probability of having clouds given the presence of rain by Eq (4.11). The term *prior probability* is given to  $P(A)$  (probability of having rain in general). The term *likelihood* is given to  $P(B|A)$  (probability of having clouds given the presence of rain). The term *posterior probability* is given to  $P(A|B)$  (probability of having rain given the presence of clouds). The term *normalization constant* is given to  $P(B)$  (probability of having clouds in general).

The posterior probability of a hypothesis H, given the data D and all other prior information I, is stated as:

$$P(H|D, I) = \frac{P(H|I)P(D|H, I)}{P(D|I)}. \quad (4.12)$$

where  $P(H|I)$  is the prior probability of H given I, or otherwise known as the prior;  $P(D|I)$  is the probability of the data given I, which is usually taken as a normalising constant;  $P(D|H, I)$  is the direct probability (or likelihood) of obtaining the data given H and I. Berger & Sellke (1987) obtained, using Bayes' theorem,  $p(H_0|\tilde{y})$  with respect to  $p(\tilde{y}|H_0)$  and  $p(\tilde{y}|H_1)$ , where  $H_1$  is the alternative hypothesis.

$$p(H_0|\tilde{y}) = \frac{p(H_0)p(\tilde{y}|H_0)}{p(H_0)p(\tilde{y}|H_0) + p(H_1)p(\tilde{y}|H_1)}. \quad (4.13)$$

I set  $p_0 = p(H_0)$ , and since we have  $p(H_1) = 1 - p_0$ , they finally obtained:

$$p(H_0|\tilde{y}) = \left( 1 + \frac{(1 - p_0)}{p_0} \mathcal{L} \right)^{-1}, \quad (4.14)$$

with  $\mathcal{L}$  being the likelihood ratio defined as:

$$\mathcal{L} = \frac{p(\tilde{y}|H_1)}{p(\tilde{y}|H_0)}. \quad (4.15)$$

Here,  $p(H_0|\tilde{y})$  is the so-called posterior probability of  $H_0$  given the observed data  $\tilde{y}$ . Naturally there is no way to favour  $H_0$  over  $H_1$ , or vice versa, otherwise our own prejudice would most likely be confirmed by the test, i.e.  $p_0 = 0.5$ . Subsequently, Berger et al. (1997) recommended to report the following when performing hypothesis testing:

$$\text{if } \mathcal{L} > 1, \text{ reject } H_0 \text{ and report } p(H_0|\tilde{y}) = \frac{1}{1 + \mathcal{L}}, \quad (4.16)$$

$$\text{if } \mathcal{L} \leq 1, \text{ accept } H_0 \text{ and report } p(H_1|\tilde{y}) = \frac{1}{1 + \mathcal{L}^{-1}}. \quad (4.17)$$

The advantage of such a presentation is that even for a borderline case, say when the ratios above are close to unity, it is clear that there is only a 50 % chance that the  $H_0$  hypothesis is wrongly accepted, or wrongly rejected. This presentation is more honest and better encapsulates human judgement and prejudice.

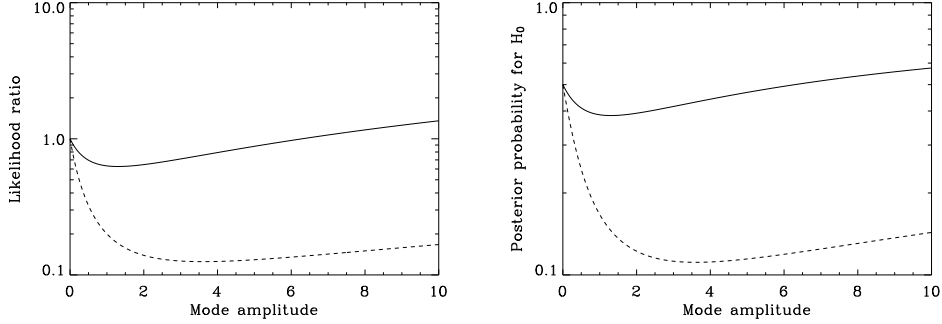


FIGURE 6. On the left-hand side, likelihood ratio  $\mathcal{L}$  as a function of the mode amplitude for *detection significances* of 10 % (solid line), and of 1 % (dashed line); the noise is set to unity. On the right-hand side, the posterior probability of  $H_0$  as a function of mode amplitude for *detection significances* of 10 % (solid line), and of 1 % (dashed line) (from Eq. 4.33)].

**Example of posterior probability.** Using the example given in Section 4.1.1 for the detection of sine waves, we can derive  $p(H_0|\tilde{y})$  using Eq. (4.6) as:

$$p(H_0|\tilde{y}) = \left( 1 + \frac{1}{1+H} p^{-H/(1+H)} \right)^{-1} \quad (4.18)$$

where  $p = e^{-\tilde{y}/B}$  is the *detection significance*. Figure 6 show the results for two different *detection significances*. When the *detection significance* is 10 %, the likelihood ratio can be greater than unity for large values of the mode amplitude, leading to the acceptance of the null hypothesis. This is rather paradoxical, i.e., that large mode amplitude can lead to the rejection of the alternative hypothesis. To resolve the paradox we note that the posterior probability of  $H_0$  is in any case never lower than 40%, or the posterior probability of  $H_1$  is never higher than 60%. This implies that both hypotheses are equally likely when the *detection significance* is as low as 10 %. In other words, when we set, a priori, a large mode amplitude and get a low *detection significance*, the alternative hypothesis is as likely as the null hypothesis. In other words, the assumption about a large mode amplitude is *not supported by the data*.

The main conclusion to be drawn from this calculation is that the *detection significance* should be set much lower than 10 % in order to avoid misinterpretation of the result. For example, with a *detection significance* of 1 %, the posterior probability for  $H_0$  can fall to 10 % when the signal-to-noise ratio is above unity. Sellke et al. (2001) showed that the posterior probability can never be lower than the lower bound:

$$p(H_0|x) \geq \left( 1 - \frac{1}{ep \ln p} \right)^{-1} \quad (4.19)$$

The reader may verify for themselves that this lower bound is effectively reached for Eq. (4.18). In the case, when the amplitude of the mode  $A$  is not known, one needs to set, a priori, the value for the likely range of amplitudes. In the case of a uniform prior, the posterior probability  $p(H_0|\hat{x})$  then does reach a minimum that is higher than the lower bound of Eq. (4.19) (Appourchaux et al. 2009).

In summary, the significance level should not be used for justifying a detection (or a non-detection). Instead I recommend using the prescription of Berger et al. (1997), as given by Eqs. (4.16) and (4.17) and to specify the alternative hypothesis  $H_1$ .

**On the choice of the prior probability** One important question when applying Bayesian statistics is what value should the prior probability of the hypothesis  $H_0$ , i.e.,  $p_0$ , take? We define the prior probability as the probability that the  $H_0$  hypothesis is correct. The probability that the alternative  $H_1$  hypothesis is correct can then be defined as  $p(H_1) = 1 - p_0$ . It is common to set  $p_0 = 0.5$  so as to avoid prejudicing one hypothesis over the other. Would we expect the probability that  $H_1$  and  $H_0$  are true to be the same in all instances? Since Bayesian statistics requires a priori knowledge, it is possible to use our knowledge of physics/astrophysics to tell us which hypothesis is more likely to be true in a given circumstance.

#### 4.2. Parameter estimation

The previous section on *hypothesis testing* is really the prerequisite when one wants to assess if there is a signal sought in the observation. Unfortunately, knowing that a signal is present does not provide any pertinent information for doing physics or astrophysics. This is the goal of *parameter estimation*.

Parameter estimation is a vast subject in statistics. Below, I will introduce the estimations that are the most commonly used in astrophysics. The estimations described hereafter are also related to the frequentist and Bayesian world. As we will see, these estimations are not so foreign from each other. The frequentist estimation can be used when the signal-to-noise ratio is high, while the Bayesian estimation is more useful (but more time consuming) when the signal-to-noise ratio is low.

##### 4.2.1. Maximum Likelihood Estimation

As shown in the *Historical Overview* section, Gauss (1809) introduced the concept of *Maximum Likelihood Estimation* or MLE. The aim of MLE is to find the set of parameters that maximise the likelihood of the observed event, this is a *point-like* estimation. Having observed a random variable  $x$  with a probability distribution  $p(x, \boldsymbol{\lambda})$ , where  $\boldsymbol{\lambda}$  is a vector of  $p$  parameters describing the model behind the random variable  $x$ , the likelihood  $L$  of  $N$  observations of  $x$  is given by:

$$L(x, N, \boldsymbol{\lambda}) = \prod_{k=1}^N f(x_k, \boldsymbol{\lambda}). \quad (4.20)$$

where the product implicitly expresses the fact that all  $x_k$  are *independent* of each other. Usually we define the logarithmic likelihood function  $\ell$  as

$$\ell(x, N, \boldsymbol{\lambda}) = \ln L(x, \boldsymbol{\lambda}) = - \sum_{k=1}^N \ln f(x_k, \boldsymbol{\lambda}). \quad (4.21)$$

The estimate of  $\boldsymbol{\lambda}$  is derived from the maximisation of the likelihood as given by Eq. (4.21) such that we have

$$\tilde{\boldsymbol{\lambda}} = \max_{\boldsymbol{\lambda}} \ell(x, N, \boldsymbol{\lambda}) \quad (4.22)$$

Such an estimator has several interesting properties in the limit of very large sample ( $N \rightarrow \infty$ ) which are:

- MLE are asymptotically unbiased,
- MLE are of minimum variance,
- MLE are asymptotically normal.

The first property implies that:

$$\lim_{N \rightarrow \infty} E(\tilde{\boldsymbol{\lambda}}) = \boldsymbol{\lambda}_0. \quad (4.23)$$

where  $\lambda_0$  is the true value of the model. The second property implies that no other asymptotically unbiased estimator has lower variance. Using the Cramer-Rao theorem (Cramer 1946; Rao 1945), this can be rewritten as:

$$\text{cov} [\tilde{\lambda}] \geq \frac{1}{I(\lambda_0)} \quad (4.24)$$

where  $I(\lambda)$  is the Fisher information matrix whose elements are given by:

$$I_{ij} = \text{E} \left[ \frac{\partial^2 \ell(x, \lambda)}{\partial \lambda_i \partial \lambda_j} \right] \quad (4.25)$$

Asymptotically we also have:

$$\lim_{N \rightarrow \infty} \text{cov} [\tilde{\lambda}] = \frac{1}{I(\lambda_0)} \quad (4.26)$$

Finally the third property, regarding the estimator being asymptotically normal, can be expressed as:

$$p(\lambda) = \mathcal{N} \left( \lambda_0, \frac{1}{I} \right) \quad (4.27)$$

where  $\mathcal{N}$  is the normal distribution. This latter equation is usually used for providing the statistical distribution of  $\tilde{\lambda}$  as:

$$p(\lambda) \approx \mathcal{N} \left( \tilde{\lambda}, \frac{1}{H(\tilde{\lambda})} \right) \quad (4.28)$$

where  $H$  is the so-called Hessian matrix whose elements are derived from:

$$H_{ij} = \left[ \frac{\partial^2 \ell(x, \tilde{\lambda})}{\partial \lambda_i \partial \lambda_j} \right] \quad (4.29)$$

with the property given by the Cramer-Rao theorem as:

$$\frac{1}{H(\tilde{\lambda})} \geq \frac{1}{I(\lambda_0)} \quad (4.30)$$

Equation (4.29) is used when computing the so-called formal error bars on  $\tilde{\lambda}$ ; as a matter of fact according to the Cramer-Rao theorem, Eq. (4.29) gives only a lower bound to the error bars.

**Significance of estimates.** When one uses Least Squares for fitting data, one can test the significance of its fitted parameters using the so-called  $R$  test (Frieden 1983). For MLE, a useful test can be used: the likelihood ratio test. This method requires maximising first the likelihood  $e^{-\ell(\omega_p)}$  of a given event where  $p$  parameters are used to described the statistical model of the event. Then if one wants to describe the same event with  $n$  additional parameters, the likelihood  $e^{-\ell(\Omega_{p+n})}$  will be maximised. The likelihood ratio test consists in making the ratio of the two likelihood. Using the logarithmic likelihood, we can define the ratio  $\Lambda$  as:

$$\ln(\Lambda) = \ell(\Omega_{p+n}) - \ell(\omega_p) \quad (4.31)$$

If  $\Lambda$  is close to 1, it means that there is no improvement in the maximised likelihood and that the additional parameters are not significant. On the other hand, if  $\Lambda \ll 1$ , it means that  $\ell(\Omega_{p+n}) \ll \ell(\omega_p)$  and that the additional parameters are very significant. In order to define a significance for the  $n$  additional parameters, we need to know the statistics of  $\ln(\Lambda)$  under the null hypothesis, i.e. when the  $n$  additional parameters are not needed to

describe the model. For this null hypothesis, Wilks (1938) showed that for large sample size the distribution of  $-2\ln\Lambda$  tends to the  $\chi^2(n)$  distribution.

**Calibration of error bars.** Before applying MLE to real data, it is always advisable to test the power of this approach on synthetic data, i.e. performing Monte-Carlo simulations. They are not merely for playing games; these simulations are real tools for understanding what we fit and how we fit it. Assuming that the statistics of the data is known, performing Monte-Carlo is useful for the following reasons:

- Assessing the model of the data
- Assessing the statistical distribution of the parameters of the model
- Assessing the precision on the fitted parameters of the models

The parameters derived by the MLE should have the desirable properties of having a normal distribution (See above); if not we advise to apply a change of variable on the fitted parameters ( $\log x$  for instance). A normal distribution is necessary to derive meaningful error bars, this is the assumption behind Eq. (4.29). In order to be able to derive a good estimate of the error bars using one realisation, the standard deviation of a large sample of fitted parameters should be equal to the mean of formal errors return by the fit, i.e. this is the approximation of Eq. (4.27) by Eq. (4.28). In other words we should have the following approximation for the inverse of the covariance of the parameters:

$$H(\tilde{\lambda}) \approx I(\lambda_0) \quad (4.32)$$

where  $H$  is related to the formal error bars and  $I$  is related to the asymptotic error bars. This calibration as expressed in this equation is key to derive meaningful error bars. I advise the reader to use such a calibration procedure for checking the formal error bars derived from software codes fitting function using Least Squares. The formal error bars derived by a single noise realisation are only a *lower limit* to the *real* error bars. This lower limit is never reached when for instance the signal-to-noise ratio is too low. In this latter case, we are very far from the asymptotic behaviour. This case is covered by the Bayesian approach to parameter estimation.

#### 4.2.2. Bayesian parameter estimation

Parameter estimation can also be done using Bayes' theorem. In this case, this is the so-called *Bayesian inference*. Using the same notation as before, I can express using Bayes' theorem the probability distribution as:

$$p(\lambda|x, I) = \frac{p(\lambda|I)p(x|\lambda, I)}{p(x|I)} \quad (4.33)$$

where  $\lambda$  are the observables for which I seek the *posterior probability*,  $x$  is the observed data set, and  $I$  is the information. The *prior probability* of the observables is given by  $p(\lambda|I)$ : this is the way to quantify our *belief* about what I seek. The likelihood is given by  $p(x|\lambda, I)$  which is exactly the  $L(x, \lambda)$  of the previous section. Therefore the *frequentist* approach is related to the *Bayesian* approach simply by the *frequentist* likelihood, the *prior* probability and the normalization factor  $p(x|I)$ . The main advantage of the Bayesian approach is that the posterior probability  $p(\lambda|x, I)$  is directly accessible while for the frequentist approach only the location of the maximum of the likelihood is known. In this latter approach, there is no direct visibility of the parameter probability distribution but only an approximation provided by Eq. (4.28). This is why the frequentist approach is a point-like estimation whereas the Bayesian approach is more global. For instance, the power of the Bayesian approach is such that it provides the full posterior probability which may not be necessarily a normal distribution but could be the sum of many normal

distribution due to many local minima. In that case, only the posterior probability can provide a correct assessment of the statistics of the derived parameters.

**Posterior probability estimation** The main difficulty in Bayesian inference is to derive the *posterior* probability. If the derivation of Eq. (4.33) is analytical then parameter estimation can easily be done (See an example in the Application to Asteroseismology section). When this is not possible, the easiest is to compute the posterior probability by using a random walk algorithm that will provide the posterior probability from a representative samples of the  $\lambda$ . A famous example of such a procedure is derived from the so-called Metropolis-Hastings algorithm (MH) (Metropolis et al. 1953; Hastings 1970). Let us see how this algorithm works in practice for a probability distribution  $p(\lambda)$  for which we want to have a representative sample. We start from a given point  $\lambda^{(t)}$  and from a known probability distribution  $Q$ . We draw at random from the probability distribution a value  $\lambda'$  knowing  $\lambda^{(t)}$  and we compute the following ratio:

$$r = \frac{p(\lambda')Q(\lambda^{(t)}|\lambda')}{p(\lambda^{(t)})Q(\lambda'|\lambda^{(t)})} \quad (4.34)$$

then the new proposed value  $\lambda'$  is accepted or rejected following this scheme:

$$\begin{aligned} \text{If } r \geq 1 & \quad \text{then } \lambda^{(t+1)} = \lambda' \\ \text{If } r < 1 & \quad \text{then} \\ & \lambda^{(t+1)} = \lambda' \text{ if } r < \alpha \\ & \lambda^{(t+1)} = \lambda^{(t)} \text{ if } r \geq \alpha \end{aligned} \quad (4.35)$$

where  $\alpha$  is a random number drawn from a uniform distribution. Asymptotically the  $\lambda^{(t)}$  will then tend to have the probability distribution  $p(\lambda)$ . The benefit of this algorithm is that the computation of the normalisation factor in the denominator of Eq. (4.33) is not needed. Another obvious benefit is that very complex probability distributions can be derived, thereby providing the potential correlations between the various parameters  $\lambda_i$ .

The difficulty in the use of the MH is not in the algorithm itself, which is quite easy to implement but in the proper choice of the input distribution  $Q$ . This is a vast subject which goes far beyond this course. The reader will find in Gregory (2005) what is required for delving into the subject of obtaining the posterior probability distribution using various techniques: Gibbs sampling, thermal annealing, convergence and so forth.

**Mean, rms and moment estimation.** As soon as the *posterior probability* is known, we can derive an estimate of the moment  $k$  of the parameter  $\lambda_i$  by deriving first the posterior probability distribution of  $\lambda_i$  only or  $p(\lambda_i|x, I)$ . This is done by integrating (or marginalising) over the so-called *nuisance* parameters as follows:

$$p(\lambda_i|x, I) = \int_{\Omega_i} p(\lambda|x, I) d\lambda_n \quad (4.36)$$

where  $\lambda_n \in \Omega_i$  with  $n \neq i$ . Then we can compute the moments of the posterior probability by writing:

$$\langle \lambda_i^k \rangle = \int_{\lambda_i} \lambda_i^k p(\lambda_i|x, I) d\lambda_i \quad (4.37)$$

The first and second moments provide the mean value and rms deviations of  $\lambda_i$  as

$$\tilde{\lambda}_i = \langle \lambda_i^1 \rangle \quad (4.38)$$

$$\sigma_{\lambda_i} = \sqrt{\langle \lambda_i^2 \rangle - \tilde{\lambda}_i^2} \quad (4.39)$$

The use of these two moments is enough to describe a normal distribution, as all the

other moments can be derived from these two. Here I note that the Cramer-Rao criteria is also relevant. It means that the rms value derived above is bounded as follows:

$$\sigma_{\lambda_i}^2 \geq H^{-1}(\tilde{\lambda}_i) \quad (4.40)$$

As a result, the error bars derived from the Bayesian approach are larger than those returned using MLE. It means that, contrary to popular belief, the Bayesian approach is more conservative than MLE.

If the posterior probability is not normal, higher moments could be quoted. It is rather impractical to quote all the moments higher than 2. Instead, we can use the value of the median and of the percentiles. We can define, for a random variable  $x$  with a probability distribution  $p$ , two values  $x_1$  and  $x_2$  such that for a given percentile  $q$  we have:

$$q = \int_{-\infty}^{x_1} p(x)dx = \int_{x_2}^{+\infty} p(x)dx \quad (4.41)$$

When  $q = 50\%$ , we have  $x_1 = x_2$  thereby providing the median. Usually for a gaussian distribution, the  $1\text{-}\sigma$  and  $2\text{-}\sigma$  values provide a percentile of 15.9% and 2.3%, respectively, i.e. 68.2% and 95.4% of the values are in the range defined by  $x_1$  and  $x_2$ . The use of the 4 percentiles and the median can be shown in a so-called box plot. The main advantage of using percentiles is that they are invariant under a change of variable. In other words, when making a change of variable from  $x$  to  $g(x)$  in Eq. (4.41), the  $x_1$  and  $x_2$  returned are unaffected by the transform (provided that  $g$  is a monotonic function of  $x$ ).

In practice, when the analytical integration cannot be done, Eq. (4.37) is computed using the MH algorithm mentioned above, such that we have:

$$\langle \lambda_i^k \rangle = \frac{1}{N_t} \sum_t \left( \lambda_i^{(t)} \right)^k \quad (4.42)$$

where  $N_t$  is the total number of samples computed returned by the MH algorithm. Then the median and the percentiles are computed by sorting the values of  $\lambda_i^{(t)}$ . Examples of such results can be found in Benomar et al. (2009a).

**Role of the prior.** The *prior probability* expresses what we believe we know (or not) about the parameters  $\lambda_i$ . The choice of the prior is related to the amount of information at our disposal. The most obvious *prior probability* of the parameters  $\lambda_i$  of interest is the one that is uniformly distributed over some range; this is an *uninformative* prior. The role of the prior and its impact on the *posterior probability* should ideally be as small as possible. Objective uninformative priors are derived using the procedure described by Jeffreys (1946), related to the calculation of the determinant of the Fisher matrix (Fisher 1925). An informative prior could be, for example, a gaussian distribution of a given parameter  $\lambda_i$ . Priors are not always *proper* in the sense that they are not always related to a *proper* probability distribution, i.e. providing finite moments. The  $1/\sigma$  prior for the unknown rms value of a parameter is a specific example of an improper prior. An extensive discussion on the impact of the prior in a Bayesian framework has been well developed by Jaynes (1987).

**Significance and model comparison.** The power of the Bayesian approach is also to be able to compare different models. The approach used by frequentists using the likelihood ratio test outlined in the previous section is quite similar to what is called the Bayesian odd ratio. Let us assume that one wants to compare between different model  $M_n$ . The odd ratios between any two sets of models is:

$$O_{n/m} = \frac{p(M_n|x, I)}{p(M_m|x, I)} = \frac{p(M_n|I)}{p(M_m|I)} \frac{p(x|M_n, I)}{p(x|M_m, I)} \quad (4.43)$$

The second part of the equation was derived from Bayes' theorem. The second fraction closely resembles the likelihood ratio presented above, but it is the product of the ratio of the *prior* model probabilities by the ratio of the *global* likelihood. It is termed *global* because these probabilities are not a point-like estimate (as in the frequentist approach) but an integration over all the possible values of the estimated parameters  $\lambda_i$ . The global likelihood  $p(x|M_m, I)$  is written as:

$$p(x|M_m, I) = \int_{\Omega} p(\lambda|M_m, I)p(x|\lambda, M_m, I)d\lambda \quad (4.44)$$

where  $p(\lambda|M_m, I)$  is the prior probability, and  $p(x|\lambda, M_m, I)$  is the likelihood. The computation of the global likelihood is rather difficult but can be done using the MH algorithm under parallel tempering. The integration of Eq. (4.44) can then be done using the so-called thermodynamic integration (Gelman & Meng 1998). Applications of this kind of integration can be found in Gregory (2005).

It is also useful to express the posterior probability of each model  $p(M_n|x, I)$  as follows:

$$p(M_n|x, I) = \frac{p(M_n|I)p(x|M_n, I)}{\sum p(M_m|I)p(x|M_m, I)} \quad (4.45)$$

Usually if the model comparison is an *objective* Bayesian analysis, then all prior model probabilities are equal ( $p(x|M_m, I) = p(x|M_k, I), \forall m, k$ ). This assumption can be used when the models are strictly different and are not *nested*. Here *nested* means that a child model relies on a parent model, the former having more parameters describing the model than the latter. Most of the models we used are indeed *nested*, they differ from each other by a few parameters. In this case, it results in the so-called model *multiplicity* that must be taken into account under the *subjective* Bayesian approach. Under this approach, it is possible to have model probability differing from each other ( $p(M_m|I) \neq p(M_k|I)$ ). A by-product of the *subjective* Bayesian approach is also to provide an estimate of these  $p(M_m|I)$  based on the data. Model multiplicity has just been started to be taken into account in model comparison (Scott & Berger 2010). Since this is very recent, I advise the reader to inform themselves on whether this approach can be useful for model comparison.

## 5. Application to asteroseismology

In the two previous sections, I laid down the foundations for applying harmonic analysis and statistics to astrophysics. In particular, the field of asteroseismology is extremely relevant for these applications. Stars have been known to oscillate since, at least, the 16<sup>th</sup> century when David Fabricius found that *o* Ceti (Mira) was variable. Since then many others stars such as  $\beta$  Cephei,  $\delta$  Scuti, Cepheids,  $\gamma$  Dor or solar-like stars have been found to oscillate (See Christensen-Dalsgaard 2004, and references therein). Stellar oscillations are mainly excited by an opacity-driven mechanism ( $\kappa$  mechanism) and by turbulence occurring in convection zones (Gautschi & Saio 1995, 1996). Broadly speaking, the world of stellar oscillations can be divided in two categories:

- Periodic pulsations having a weakly time-dependent amplitude
- Oscillatory eigenmodes stochastically excited

The first type results from overstable oscillations in stars being driven by the  $\kappa$  mechanism, whose amplitudes are limited by a non-linear mechanism. The functional form of the variation of the luminosity with time is then periodic but not necessarily sinusoidal. In that case the amplitude of the oscillations varies more slowly than the periods of the oscillations.

The second type results from modes being randomly excited in solar-like stars whose amplitudes are damped by various mechanisms (Houdek et al. 1999). In that case, the amplitude of the oscillations varies on time scales shorter than the periods of oscillations. There are stars for which the two types of oscillations co-exists (Belkacem et al. 2009, 2010).

When applying various tools for obtaining the frequencies of the oscillation (frequencies which describe the internal structure of the star), then one should ask oneself which type of functional form the stellar oscillation will have. Typically there are three type of functional forms:

- Periodic non-sinusoidal
- Sinusoidal
- Harmonic oscillator stochastically excited

These functional forms being periodic the Fourier transform is the obvious choice for time series analysis. As for the last functional form, the random nature of the excitation imposes the application of a proper statistical treatment of the time series and its associated power spectrum. Hereafter, I will treat the two most common cases encountered in asteroseismology: Classical pulsators (periodic non-sinusoidal function), Solar-like oscillators (harmonic oscillators)

### 5.1. *Classical pulsators*

For stars having luminosity variations whose functional forms is sinusoidal, the common practice is to use the Fourier transform described in the previous sections. The first application of statistics and Fourier transform was for finding periodicities in earthquakes due to Schuster (1897), which has since then be coined the *Schuster periodogram*.

For variable stars (or classical pulsators), the first application of the periodogram is attributed to Wehlau & Leung (1964). Since that date, the analysis of time series has been evolving to take into account various aspects related to the presence of gaps and the large dynamic range in the amplitude of the periodicities. One of the problem encountered when observing stars from the ground is that the periodic gaps in the observation (due to the day-night cycle) introduce aliases. The presence of these gaps produce then spurious peaks located on either side of the main peak located at multiple of  $\pm 11.57 \mu\text{ Hz}$  ( $1/24\text{ hr}$ ). One solution for taking into account such a frequency response is to apply the CLEAN algorithm (Roberts et al. 1987) which is used in radioastronomy for aperture synthesis (Högbom 1974). Another approach is to use a combinaison of the CLEAN algorithm and of prewhitening which consists in removing the signal of largest amplitude in the time series (after having being detected) and then to re-compute the periodogram; the procedure iterates until there is no large signal detected in the periodogram (Belmonte et al. 1991). This latter technique is now the most commonly used for classical pulsators.

#### 5.1.1. *Spectral analysis revisited.*

**Single sine wave.** I already touched upon the Fourier analysis of pure sine waves using the frequentist framework (application of Least Squares). Here I shall briefly revisit what can be done when using a Bayesian approach. Bretthorst (1988), using a Bayesian approach to the analysis of the time series of a pure sine wave sampled regularly and embedded in noise having a gaussian distribution, demonstrated that the *posterior probability* for the frequency  $\nu$  of the sine wave can be written as:

$$\ln P(\nu|x, \sigma, I) = \frac{C(\nu)}{\sigma^2} + \dots \quad (5.1)$$

where  $x$  are the data ( $x_i$  taken at time  $t_i$ ),  $\sigma$  is the rms value of the noise assumed to be known, and  $C(\nu)$  is the Schuster periodogram given by:

$$C(\nu) = \frac{1}{N} \left| \sum_{i=1}^N d_i e^{i2\pi\nu t_i} \right|^2 \quad (5.2)$$

This is nothing less than the power spectrum. For a pure sine wave with frequency  $\nu_0$ , the periodogram has a maximum at that frequency. Using the posterior probability for  $\nu$ , we then have the first moment of the frequency as:

$$\tilde{\nu} = \int P(\nu|x, \sigma, I) d\nu = \nu_0 \quad (5.3)$$

Following this approach, Jaynes (1987) demonstrated using Eq. (5.2) that for a pure sine wave of amplitude  $A$ , the rms error on the frequency  $\nu_0$  is given by:

$$\delta\nu = \frac{\sqrt{6}}{\pi} \frac{\sigma}{A} \frac{\sqrt{\Delta t}}{T^{\frac{3}{2}}} \quad (5.4)$$

where  $T$  is the observing time and  $\Delta t$  is the sampling time. This formula is the same as given by Cuypers (1987) and derived by Koen (1999). The frequency precision is inversely proportional to the signal-to-noise ratio computed in the time domain. This equation also shows that the Rayleigh criterion ( $1/T$ ) for frequency resolution is very pessimistic compared to the precision with which frequencies can be measured.

**Many sine waves.** The approach outlined above linking the posterior probability to the Fourier transform justifies *a posteriori* the use of that transform: "*The highest peak in the discrete Fourier transform is an optimal frequency estimator for a data set which contains a single harmonic frequency in the presence of Gaussian white noise*", Bretthorst (1988). Fortunately, for several harmonic signals whose frequencies ( $\nu_i$ ) are far from each other, the *naive* use of Fourier can be broken down in the application of the preceding section as many times as required or:

$$\ln P(\nu_1, \dots, \nu_r | x, \sigma, I) = \left( \sum_{j=1}^r \frac{C(\nu_j)}{\sigma^2} \right) + \dots \quad (5.5)$$

This justifies the use of the periodogram for finding harmonic signal in a time series (Bretthorst 1988).

**Periodic signals.** When the signal is periodic but not sinusoidal, the decomposition using the periodogram will provide the fundamental  $\nu_f$  of the period and spurious frequencies at  $n\nu_f$ . The presence of these spurious frequencies can become difficult to handle when there are a lot of frequencies such as in  $\delta$  Scuti (Poretti et al. 2009). The periodic signal can also be the result of non-linear saturation which cannot be easily described in terms of harmonic signals (Yoachim et al. 2009). As mentioned above, prewhitening together with the use of the CLEAN algorithm is a possible solution for analysing such periodic signals. The analysis can also be performed using other techniques such as Principal Components Analysis (Tanvir et al. 2005).

I showed in a previous section how one could analyse data, that are not equally sampled in time, using the Lomb-Scargle periodogram. The LS periodogram has been extended not only to a decaying sinusoid (Bretthorst 2001a) but also to periodic functions in general (Bretthorst 2001b). This revisitation of spectral analysis by Bretthorst is extremely useful when one wants to understand how to apply a Bayesian analysis to time series. Based on the work of Bretthorst (2001b), a possible application would be to model the

amplitude limitation  $\delta$  Scuti to obtain the functional forms of the periodic signal. A different functional form from a sinusoid could be used as an advantage for disentangling the various possible combinations of frequencies (Poretti et al. 2009) either due to the gaps, the periodic signal or the physics of the stars. This is an avenue yet to be explored.

## 5.2. Solar-like oscillations

The analysis of solar-like oscillations is slightly more complicated because of the random excitation of the modes. In order to apply the Fourier transform and use statistical tools for extracting mode frequencies and mode parameters for such stars, one has to understand how the functional form of the mode amplitude is related to the harmonic oscillator being randomly excited.

### 5.2.1. Randomly excited harmonic oscillations

It is well known that pressure modes (or p modes) are stochastically excited oscillators (Kumar et al. 1988). The source of excitation lies in the many granules covering the star (Turbulent convection, Houdek et al. 1999; Samadi & Goupil 2001). The modes are assumed to be independently excited (Kumar et al. 1988) because the eigenfunction of the modes is primarily radial and nearly independent of degree in the upper convection zone, where the modes are excited. The eigenmodes are harmonic oscillators being intrinsically damped, and excited through a forcing function  $F$ . The differential equation of such a damped harmonic oscillator is written as:

$$\frac{d^2 x_{\text{osc}}}{dt^2} + 2\pi\gamma \frac{dx_{\text{osc}}}{dt} + (2\pi)^2 \nu_0^2 x_{\text{osc}} = F(t) \quad (5.6)$$

where  $t$  is the time,  $x_{\text{osc}}$  is the displacement,  $\gamma$  is the damping term related to the mode linewidth,  $\nu_0$  is the frequency of the mode and  $F(t)$  is the forcing function. Equation (5.6) is also the expression of an auto-regressive (AR) process which in this case is a stationary process. From this equation the Fourier transform of  $x$  can be written as:

$$\tilde{x}_{\text{osc}}(\nu) = \frac{\tilde{F}(\nu)}{(2\pi)^2(\nu_0^2 - \nu^2 + i\gamma\nu)} \quad (5.7)$$

where  $\tilde{x}_{\text{osc}}(\nu)$  and  $\tilde{F}(\nu)$  are the Fourier transform of  $x_{\text{osc}}(t)$  and  $F(t)$ . From the large number of granules, it can be derived that the forcing function is normally distributed. Therefore the 2 components (the real and imaginary parts) of the Fourier transform of the forcing function are also normally distributed (See Section 3.7). Therefore, for the harmonic oscillator, each component of  $\tilde{x}_{\text{osc}}(\nu)$  is normally distributed with a mean of zero, and the same variance. The variance is related to the spectral density given by:

$$S_{\text{osc}}(\nu) = \frac{S_F(\nu)}{(2\pi)^4[(\nu_0^2 - \nu^2)^2 + \nu^2\gamma^2]} \quad (5.8)$$

The spectral density of  $\tilde{x}_{\text{osc}}(\nu)$  has then a  $\chi^2$  with 2 degrees of freedom statistics. For such statistics, I outline that the variance is the same as the mean. Equation (5.8) is the p-mode profile that is usually approximated by a Lorentzian profile when  $\nu \approx \nu_0$  as:

$$S_{\text{osc}}(\nu) \approx \frac{S_F(\nu)}{\gamma^2(2\pi)^4\nu_0^2} \frac{1}{1+x^2} \quad (5.9)$$

with  $x = 2(\nu - \nu_0)/\gamma$ . An asymmetry effect can also be introduced in the profile of Eq. (5.9). The asymmetry effect was first detected with instruments making an image of the Sun (Duvall et al. 1993). The asymmetry is due to the location of the excitation of the modes with respect to the resonant cavity. There is a direct analogy with

a source being placed inside a Fabry-Perot cavity (Duvall et al. 1993). The asymmetry was then detected by Toutain et al. (1997) with instruments observing the Sun as a star; the sign of the asymmetry depending upon the observables (intensity of velocity). The empirical theoretical explanation for the different asymmetry in intensity with respect to velocity was given by Nigam et al. (1998). It is related to the different correlation between the background and the modes in velocity and in intensity which were studied by Severino et al. (2001). The theoretical framework provided then a simple expression for the expression of the mode profile with asymmetry (Nigam & Kosovichev 1998), given by:

$$S_{\text{osc}}(\nu) = H \frac{1 + 2Bx}{1 + x^2} + B^2 \quad (5.10)$$

where  $B$  is the asymmetry effect and  $H$  is the mode height. The effect of the linear slope is to change the sign of asymmetry when the sign of  $B$  changes. When  $B$  is positive (negative) there is more power at high (low) frequency. Therefore, if the asymmetry is not taken into account, the sign of  $B$  will affect the true location of the mode frequency differently. This effect, if not taken into account, may provide large systematic errors when comparing different observables thereby providing potential problems for the inferred model of the star.

### 5.3. *Random non-harmonic field*

Another source contributing to the observed power spectrum is the noise generated by the star itself. It is quite customary to have the stellar noise increasing at low frequencies. As a matter of fact, the noise observed is not limited to stars but more related to an intrinsic behaviour encountered in many physical phenomena. This so-called  $1/f$  noise is so ubiquitous that it is found in almost any physical measurements and also in stars. The  $1/f$  noise appears in many electrical applications and other applications (Keshner 1982). While white noise is known not to have any memory (autocovariance is the Dirac distribution), the  $1/f$  noise possesses some memory. Mathematically, AR models have also these properties. Random processes following AR models have the desired properties of having memory, thereby producing  $1/f$ -like power spectra. This fact was used by Harvey (1985) for deriving the solar background noise spectrum for instruments observing solar radial velocities. The spectrum is derived from a superposition of four components related to solar activity and to three different types of granulation, having different lifetimes. In Harvey (1985), the power law used had a -2 slope; it was later refined to be arbitrary  $b$  (Harvey et al. 1993) such that we have for the spectral density of background noise  $x_n(t)$ :

$$S_n(\nu) = \sum_i \frac{H_i}{1 + (2\pi\tau_i\nu)^{b_i}} \quad (5.11)$$

where  $\tau_i$  is the characteristic time of the decaying autocorrelation function of the process,  $H_i$  is the height in the power spectrum at  $\nu = 0$ , and  $b_i$  is the power law exponent (with the  $b_i$  ranging typically from 2 to 6 for intensity measurements; Appourchaux et al. 2002; Aigrain et al. 2004). It is clear that Eq. (5.11) also has a Lorentzian-like shape as much as the harmonic oscillator of the previous section. Therefore I also outline that modes stochastically excited since they also follow an AR model, have also the same mathematical property as the random non-harmonic field.

### 5.4. *Power spectrum model*

**Stationary processes.** I showed in Section 3.7 that stationary processes have indeed a  $\chi^2$  with 2 d.o.f statistics. The question is: what is the mean value when different stationary

processes are in play? Let us assume that two time series for two different processes co-exist:  $x_{\text{osc}}(t)$  related to the excitation of the harmonic oscillators (the eigenmodes),  $x_n(t)$  related to the non-harmonic noise (the background stellar noise). The total time series is then given by:

$$x(t) = x_{\text{osc}}(t) + x_n(t) \quad (5.12)$$

and the autocorrelation is given by:

$$\text{E}[x(t_1)x(t_2)] = C_{\text{osc}}(\tau) + C_n(\tau) + [\text{E}[x_{\text{osc}}(t_1)x_n(t_2)] + \text{E}[x_n(t_1)x_{\text{osc}}(t_2)]] \quad (5.13)$$

with  $\tau = t_2 - t_1$  and where  $C_{\text{osc}}$ ,  $C_n$  are the autocorrelation for the harmonic and non-harmonic signals. The last term in brackets is related to the potential correlation between the two type of stationary processes. Usually, we assume that there is no correlation between these processes such that the term in brackets is zero. Using the Wiener-Khinchine theorem, we have for the spectral density:

$$S_x(\nu) = S_{\text{osc}}(\nu) + S_n(\nu) \quad (5.14)$$

So the spectral density of the sum of two independent stationary process is the sum of the spectral density of each stationary process. This implicit assumption, not frequently mentioned, is usually a good approximation for solar-like oscillations. If we assume that the processes are correlated but that their correlation is stationary we can then write:

$$S_x(\nu) = S_{\text{osc}}(\nu) + S_n(\nu) + \tilde{I}(\nu) \quad (5.15)$$

where  $\tilde{I}(\nu)$  is simply the Fourier transform of the expression in brackets of Eq. (5.13). Such correlations between the background and the harmonic oscillators have been studied by Severino et al. (2001). The term  $\tilde{I}$  is responsible for the mode profile asymmetry as explained above (Nigam & Kosovichev 1998).

**Model of the spectrum.** The complete model of the power spectrum is given by the superposition of *all* the harmonic signals and of the non-harmonic signals. Apart from the Sun, so far we have only observed disk-integrated oscillations in stars. Instruments integrating over the stellar surface observe the velocity or the intensity signal as a superposition of various modes of different degrees. The impact of this integration is to filter out the higher degree modes, keeping typically only the degrees below 4. The resulting mode sensitivity ( $V_l$ ) was first computed for velocity measurements for non-rotating stars by Christensen-Dalsgaard & Gough (1982), then by Christensen-Dalsgaard (1989) taking into account Doppler imaging due to rotation; for intensity, it was first computed by Toutain & Gouttebroze (1993). The impact of the inclination angle of the star upon the mode sensitivity ( $c_{l,m}$ ) was also computed by Toutain & Gouttebroze (1993), calculation which was later re-discovered by Gizon & Solanki (2003). The mode sensitivity  $V_l$  has also been extensively computed by different authors; e.g. Bedding et al. (1996), Appourchaux et al. (2000) and Ballot et al. (2006).

Using the equations given above and taking into account the geometrical mode sensitivity, the full spectral density is therefore given by:

$$S_x(\nu) = \sum_{n,l,m} H_n V_l^2 c_{l,m} \tilde{L} \left( \frac{\nu - \nu_{nlm}}{\Gamma_{nlm}} \right) + \sum_i \frac{H_i}{1 + (2\pi\tau_i\nu)^{b_i}} \quad (5.16)$$

where  $H_n$  is the mode height for radial order  $n$ ,  $\tilde{L}$  is the mode profile,  $\Gamma_{nlm}$  is the inverse mode lifetime,  $\nu_{nlm}$  is the mode frequency; and the  $H_i$ ,  $\tau_i$  and  $b_i$  represent the stellar and instrumental background noise (including also the photon noise). The mode profile is here assumed to be Lorentzian. It can be replaced by an asymmetrical profile but at the expense of making an incorrect approximation of the correlation effect  $\tilde{I}$ . If one wants

to properly take into account this correlation effect, one should refer to Severino et al. (2001).

Several effects will lift the degeneracy of the mode frequency  $\nu_{nl}$  such as stellar rotation or stellar activity, thereby splitting the frequency of the mode in different component (Zeeman-like effect). In this case, the mode frequency can be decomposed into the Clebsch-Gordan coefficient as:

$$\nu_{nlm} = \nu_{nl} + \sum_{i=1}^{i_{\max}} a_i(n, l) l \mathcal{P}_l^{(i)}(m/l) \quad (5.17)$$

where  $i_{\max} \leq 2l$ , and  $a_i(n, l)$  are the splitting coefficients and  $\mathcal{P}_l^{(i)}$  are polynomials given by Ritzwoller & Lavelly (1991) or by Chaplin et al. (2004). The advantage of the  $\mathcal{P}_l^{(i)}$  is that they are by construction orthogonal to each other. With this property, the max  $i$  for the decomposition can be different from  $2l$ , in other words removing or adding polynomials will not affect the value of the  $a_i(n, l)$ . To second order, this property may not be completely true as there could be slight variations in the mode height due to stochastic excitation or other effects not modelled. For example, the observation of the integrated signal over the stellar surface with a star observed perpendicular to its rotation axis ( $i = 90^\circ$ ) will *remove* modes for which  $l + m$  is odd. If this averaging effect is not taken into account the decomposition will affect the value of the  $a_i(n, l)$  (Chaplin et al. 2004).

Depending on the model used, different assumptions concerning the mode and background parameters are possible. For instance, the mode linewidth can be assumed to be independent of frequency, or to vary with frequency according to some polynomials, or to be independent for each order  $n$ , or to be independent for each degree  $l$ , and so forth and so on. It is sometimes desirable to reduce the number of fitted parameters using such assumptions. Examples of such assumptions are found in the references given in the next two sections.

### 5.5. *Frequentist parameter estimation*

Now having observed a star, we have a time series of intensity or velocity fluctuations which is properly sampled, and for which the filtering effect of the data reduction is understood. We compute the Fourier transform and then obtain the power spectral density properly scaled using Parseval's theorem. The amount of gaps is negligible such that the frequency bins are independent of each other. The statistics of the power spectrum is  $\chi^2$  with 2 d.o.f. Assuming that all the stationary processes are nearly independent, we have a complete model of the power spectrum including harmonic and non-harmonic signals.

Now that I have set the scene, all the actors are in place for the play. The method applied for deriving all the parameters of Eq. (5.16) is usually to use MLE. The use of MLE for fitting solar power spectrum was first mentioned by Duvall & Harvey (1986), then applied and tested by Anderson et al. (1990); in which they also mentioned the use of Monte-Carlo simulations for validating the method. Their pioneering work led in helioseismology to the understanding of the derivation of error bar for mode frequencies (Libbrecht 1992; Toutain & Appourchaux 1994) which is given by:

$$\delta\nu = \sqrt{\frac{\Gamma}{4\pi T}} (\beta + 1)^{\frac{1}{4}} \left( \sqrt{\beta + 1} + \sqrt{\beta} \right)^{\frac{3}{2}} \quad (5.18)$$

This equation shows that the mode precision is proportional to the square root of the mode linewidth and is proportional to the square root of the frequency resolution. This results greatly contrasts with that of a pure sine wave given by Eq. (5.4).

Similarly other error bars were also derived for linewidth, mode height, white noise (Toutain & Appourchaux 1994); rotational splitting (Toutain & Appourchaux 1994; Gizon & Solanki 2003; Ballot et al. 2008); inclination angle (Gizon & Solanki 2003; Ballot et al. 2008); and of the correlation between the mode height, the linewidth and background noise (Toutain & Appourchaux 1994) (See Appendix A); and between the rotational splitting and the inclination angle (Ballot et al. 2008).

Monte-Carlo simulations have also been used for verifying the veracity of using the inverse of the Hessian for estimating error bars on the parameters of the model (Anderson et al. 1990; Toutain & Appourchaux 1994; Schou & Brown 1994; Appourchaux et al. 1998; Gizon & Solanki 2003) (See Section 4.2.1). When doing so, it was patent that the distribution for mode height, linewidth and background noise was not normal but log-normal. There are two main reasons for this state of affairs: first, these simulations are far from the asymptotic behaviour, otherwise the distribution would be normal; second, the mathematical transform  $\log(x)$  allows us to have a distribution closer to a normal distribution than  $x$  alone. In other words, the asymptotic behaviour is reached faster with such a transform. The same would apply to any transform reducing the amount of value larger than the median with respect to value smaller than the median<sup>†</sup>.

Although the MLE has been in use for more than 20 years by helioseismologists, it is not yet completely adopted by our fellow asteroseismologists. The body of evidence provided by helioseismology is rather large, hereafter I will necessarily refer to a few examples: IPHIR (InterPlanetary Helioseismology by IRradiance measurements) photometers aboard the Russian mission Mars94 (Sun as a star; Toutain & Fröhlich 1992), VIRGO (Variability of solar IRradiance and Gravity Oscillations) photometers aboard SoHO (Solar and Heliospheric Observatory) (Sun as a star and low resolution; Fröhlich et al. 1997; Appourchaux et al. 1997), GOLF (Global Oscillations at Low Frequency) spectrometer aboard SOHO (Sun as a star, Gelly et al. 2002), BiSON (Birmingham Solar Oscillations Network) instrument (Sun as a star Chaplin et al. 1996), LOWL (Low- $l$ ) instrument (imager; Schou 1992; Schou & Brown 1994), GONG (Global Oscillation Network Group) instrument (imager; Hill et al. 1996; Appourchaux et al. 1998), SOI/MDI (Solar Oscillations Investigation / Michelson Doppler Imager) instrument aboard SoHO (imager; Kosovichev et al. 1997).

Following these successful applications for solar data, MLE was preliminarily tested on synthetic stellar timeseries for use on the CNES CoRoT mission (Appourchaux et al. 2006a,b). The recipe laid down in this latter publication was found later not be applicable to the real CoRoT data. Appourchaux et al. (2008) introduced a scheme using not fit on narrow frequency window (*local* fit as in Appourchaux et al. 2006a) but using fit on large frequency window (*global* fit). This new scheme has now been applied to solar-like stars and red giants observed by CoRoT (García et al. 2009; Barban et al. 2009; Mosser et al. 2009; Deheuvels et al. 2010; Baudin et al. 2011); and to solar-like stars observed by the NASA (National and Aeronautics Space Administration) Kepler mission (Mathur et al. 2011).

One of the recent controversies in asteroseismology is related to the degree identification in the power spectra of HD49333 (observed by CoRoT). In Appourchaux et al. (2008) the degree of the modes could not easily identified leaving two possible solutions: the ridges in the echelle diagram were either  $l = 0 - 1$  or  $l = 1 - 0$ . The reason for such a difficult identification, already anticipated by Appourchaux et al. (2006a), is related to a mode linewidth larger than in the Sun combined with a narrower small spacing than

<sup>†</sup> The reader may convince themselves by simulating a  $\chi^2$  with  $n$  d.o.f for example together with a  $\log x$  or  $\sqrt{x}$  transform.

in the Sun. As a result, the two ridges  $l = 0 - 2$  and  $l = 1$  were indistinguishable from each other. One of the solutions found by Appourchaux et al. (2008) was to base the choice of the ridge upon the likelihood ratio test. Here I would like to cite a warning in Appourchaux et al. (2008) as to the use of this test: "*However, it is important to remember that a higher likelihood does not mean that a given model is physically more meaningful, rather, it means the model is statistically more likely.*" It was clear at this stage that in order to take into account prejudices and a priori knowledge that could have helped to get the physically meaningful model, a Bayesian approach to the problem was required (ibid). This is the subject of the next section.

### 5.6. Bayesian parameter estimation

In asteroseismology, the first attempt at applying a Bayesian approach was due to Brewer et al. (2007). They used a model based upon pure sine waves which did not correctly model stochastic excitation of harmonic oscillations. Their model can be used for classical pulsators or when the modes have a lifetime longer than the observation time. Following the attempt by Brewer et al. (2007), I realised that the approach and framework provided by Bretthorst (1988) is not generally applicable in asteroseismology. The separation in time between the deterministic signal and the noise cannot be done when oscillators are stochastically excited. Since the separation is not possible, the statistics of the stellar signal suggested by Brewer et al. (2007) needs to be replaced by the likelihood usually used by frequentist Appourchaux (2008). At the time, it was not clear that the replacement was so *obvious*.

Following the early work of Appourchaux (2008), the application of Bayesian analysis to solar-like oscillators has been quite extensively developed. The difficult data analysis of HD49333 led to the development of several Bayesian data analyses in asteroseismology (Benomar et al. 2009a; Gruberbauer et al. 2009). A simplified Bayesian approach called Maximum A Posteriori (MAP) has also been used for putting constraints on the mode height (Gaulme et al. 2009). Recently, Handberg & Campante (2011) produced a very nice guide for anyone wanting to do Bayesian data analysis in asteroseismology. A more extensive guide in French is also available in Benomar (2010). These two guides covers many aspect of the application of Bayes' theorem such as power spectrum models, statistics, priors, parallel tempering and global likelihood. Here I should add that the Bayesian framework has been used not only for parameter estimation but also for making decision based upon posterior probability estimates: mode detection (Appourchaux et al. 2009; Broomhall et al. 2010), presence of  $l = 3$  modes and mixed modes (Deheuvels et al. 2010).

As anticipated by Appourchaux et al. (2008), the HD49333 controversy led to the first useful application of the Bayesian approach. Benomar et al. (2009a) using the same time series as Appourchaux et al. (2008) applied a Bayesian analysis and derived the global likelihood of the various models previously used. In doing so, Benomar et al. (2009a) showed that amongst the four models used the identification of Appourchaux et al. (2008) was the most probable one *but* for a model having *no*  $l = 2$  modes ! When these latter are included (as in Appourchaux et al. 2008), Benomar et al. (2009a) found that the identification of Appourchaux et al. (2008) was less probable (14%) than the other identification (23%); thereby showing that both models were nearly equiprobable. The absence of  $l = 2$  modes in HD49333 was also the conclusion of Gruberbauer et al. (2009) who also used a Bayesian approach. The main differences of the work of Gruberbauer et al. (2009) from that of Benomar et al. (2009a) are that they do not use global likelihoods for decision and that their model has no splitting (no degree identification possible). The controversy was finally closed when 180 days of data were available for HD49333.

Using these data, Benomar et al. (2009b) showed that both frequentist and Bayesian approaches favour a model having  $l = 2$  modes with a degree identification different from that of Appourchaux et al. (2008).

What lessons can we draw from such a controversy? The process of the advancement of science cannot be done without making *errors*. The can be of several kind:

- Done on purpose for hiding the facts (a lie)
- Done by ignorance of the facts
- Done in knowledge of the facts

The first kind of error has lead to several infamous publication such as a cold fusion (Fleischmann & Pons 1989) finding which was never reproduced. The second kind of error is due to lack of information or knowledge; in this world of information flow this can easily happen to anyone of us. The third kind can either be perceived as an error from an outsider or as a decision done in full conscience of what has been decided. The *error* of Appourchaux et al. (2008) is obviously akin to the last kind. If we would have had the choice, we would have certainly opted for applying the Bayesian approach in Appourchaux (2008) but time prevented us to do so. This is another lesson to be learnt: even in this fast paced world of information flow and publications, one should never trade quality of publication and scientific integrity for fast publication. Science is like pasta, a slow sugar.

## 6. Conclusion

If you have been able to survive until now, I would like to thank you. It would be pretentious to believe that this *Conclusion* is going to conclude anything for good. As a matter of fact, this conclusion aims at giving the reader the incentive to go beyond this *course* but still using this *course* as a foundation for the future. Data analysis clearly starts from the instrumentation itself, this aspect was not treated here. It is clear that the measurement leading to the analysis of time series is done through an instrument. From this point of view, the aspects related to the filtering due to the instrument are implicitly laid out. Apart from these mere details, the reader should find all that is needed for a proper understanding of data analysis in asteroseismology. Needless to say, that this field is evolving very rapidly, thanks to the CoRoT and Kepler missions. In the last couple of years, the Bayesian data analysis has developed quickly. Despite what is commonly thought, the Bayesian approach although using prejudices is in the end more conservative than the frequentist approach. Actually, the main drawback of this approach is not in the methodology itself but the speed at which the computation of the results can be performed.

With the advent of the new PLATO mission, it can be anticipated that both Bayesian and frequentist data analysis will be used for deriving the mode parameters of more than 20,000 stars. I can think of a hybrid scheme where both approaches will be used depending on the signal-to-noise ratio in the power spectrum. Another aspect which was partly discussed here is how one can assess and insure that the fitted mode parameters are of usable quality for doing stellar models. This is certainly to be one of the future challenges of asteroseismology: the automatic assessment of stellar mode parameters. I can also anticipate that we will apply quality assessment which are regularly used in factories doing mass production. For instance, sampling at random some of our output products may be useful for *qualifying* the lot as useable by the scientific community. We are really in the infant stage for such massive quality assessment.

The future of asteroseismology will reach another milestones when the surface of stars will be imaged. At that point time, we will then be able to probe the internal and

dynamics structure of stars in the very same manner as we did with the Sun. The Stellar Imager could be the next asteroseismic challenge whose achievement may be not so far in the future (Christensen-Dalsgaard et al. 2011).

As for any new endeavour, we need to learn from the past for going ahead while keeping in mind Timothy Leary's motto: *Think for yourself, question authority.*

**Acknowledgments.** *I would like to thank Pere Pallé for inviting me to give these lectures in the beautiful island of Tenerife. I would also like to express my thanks for the many discussions I had with my long standing colleagues from which this course benefited: Frédéric Baudin, Othman Benomar, Patrick Boumier, William Chaplin, Patrick Gaulme, Laurent Gizon and Takashi Sekii. This course also benefited from the acute reading of John Leibacher. Last but not least, many thanks to my wife, Maryse, and my sons, Kévin and Thibault, for their indefectible support; without forgetting the mention of the purring cat, Myrtille. Thanks to the volcano of El Teide for momentary time support, going up there was not an easy ride !*

## Appendix A. Correlation between mode height and linewidth

Toutain & Appourchaux (1994) showed that some fitted mode parameter were intrinsically correlated. When one wants to derive the mode amplitude  $A$  (proportional to the square root of  $\Gamma H$ ) one should take into account the correlation between the errors on  $\Gamma$  and  $H$  which are the mode linewidth the mode height, respectively. Since we have  $A \propto \Gamma H$ , and using the logarithm of these quantities, we have:

$$\log A = \frac{1}{2}(\log \Gamma + \log H) + \dots \quad (\text{A } 1)$$

or simply

$$a = \frac{1}{2}(\gamma + h) + \dots \quad (\text{A } 2)$$

The error bar on  $a$  is derived from:

$$\sigma_a^2 = \frac{1}{4}(\sigma_\gamma^2 + \sigma_h^2 + 2E(ah)) \quad (\text{A } 3)$$

which can be rewritten using the correlation coefficient as:

$$\sigma_a^2 = \frac{1}{4}(\sigma_\gamma^2 + \sigma_h^2 + 2\rho\sigma_\gamma\sigma_h) \quad (\text{A } 4)$$

Using the work of Toutain & Appourchaux (1994), and after some mathematical manipulation, I can derive for a single mode the correlation as:

$$\rho = \frac{E(\gamma h)}{\sigma_h\sigma_\gamma} = -\sqrt{\frac{\sqrt{\beta+1}}{\sqrt{\beta+1} + \sqrt{\beta}}} \quad (\text{A } 5)$$

where  $\beta$  is the noise to the mode height ratio in the power spectrum. This correlation is negative, and its absolute value always greater than 76.5% when  $\beta < 1$ . It means that even when the modes are easily detected the correlation is always very large and never negligible. Finally the error bar on  $a$  is then given by:

$$\sigma_a = \frac{1}{2\sqrt{T}\Gamma\pi} \sqrt{\left(\sqrt{\beta+1} + \sqrt{\beta}\right) \left(\left(\sqrt{\beta+1} + \sqrt{\beta}\right)^3 - 4(1+\beta)\sqrt{\beta}\right)} \quad (\text{A } 6)$$

where  $T$  is the observation time. From this equation, I can deduce that the precision

on the measurement of the mode amplitude increases with the observation time and the mode linewidth.

## REFERENCES

- Aigrain, S., Favata, F., & Gilmore, G. 2004, *A&A*, **414**, 1139
- Anderson, E. R., Duvall, T. L., J., & Jefferies, S. M. 1990, *ApJ*, **364**, 699
- Appourchaux, T. 2008, *Astronomische Nachrichten*, **329**, 485
- Appourchaux, T., Andersen, B., & Sekii, T. 2002, in *ESA Special Publication*, Vol. 508, From Solar Min to Max: Half a Solar Cycle with SOHO, ed. C. Fröhlich and A. Wilson, 47–50
- Appourchaux, T., Andersen, B. N., Fröhlich, C., et al. 1997, *Sol. Phys.*, **170**, 27
- Appourchaux, T., Berthomieu, G., Michel, E., et al. 2006a, Data analysis tools for the seismology programme (The CoRoT Mission (Eds) M. Fridlung, A. Baglin, J. Lochard and L. Conroy, ESA Publications Division, ESA Spec. Publ. 1306), 377
- Appourchaux, T., Berthomieu, G., Michel, E., et al. 2006b, Evaluation of the scientific performances for the seismology programme (The CoRoT Mission (Eds) M. Fridlung, A. Baglin, J. Lochard and L. Conroy, ESA Publications Division, ESA Spec. Publ. 1306), 429
- Appourchaux, T., Fröhlich, C., Andersen, B. N., et al. 2000, *ApJ*, **538**, 401
- Appourchaux, T., Gizon, L., & Rabello-Soares, M. C. 1998, *A&A Sup. Series*, **132**, 107
- Appourchaux, T., Michel, E., Auvergne, M., et al. 2008, *A&A*, **488**, 705
- Appourchaux, T., Samadi, R., & Dupret, M. 2009, *A&A*, **506**, 1
- Ballot, J., Appourchaux, T., Toutain, T., & Guittet, M. 2008, *A&A*, **486**, 867
- Ballot, J., García, R. A., & Lambert, P. 2006, *MNRAS*, **369**, 1281
- Barban, C., Deheuvels, S., Baudin, F., et al. 2009, *A&A*, **506**, 51
- Baudin, F., Barban, C., Belkacem, K., et al. 2011, *ArXiv e-prints*, 1102.1896
- Bayes, T. 1763, *Philosophical Transactions of the Royal Society of London*, **53**, 370
- Bedding, T. R., Kjeldsen, H., Reetz, J., & Barbuy, B. 1996, *MNRAS*, **280**, 1155
- Belkacem, K., Dupret, M. A., & Noels, A. 2010, *A&A*, **510**, A6
- Belkacem, K., Samadi, R., Goupil, M., et al. 2009, *Science*, **324**, 1540
- Belmonte, J. A., Chevreton, M., Mangeney, A., et al. 1991, *A&A*, **246**, 71
- Benomar, O. 2010, PhD thesis, Université Paris-Sud XI, Orsay, France
- Benomar, O., Appourchaux, T., & Baudin, F. 2009a, *A&A*, **506**, 15
- Benomar, O., Baudin, F., Campante, T. L., et al. 2009b, *A&A*, **507**, L13
- Berger, J. 1997, *Statistical Science*, **18**, 1
- Berger, J., Boukai, B., & Wang, Y. 1997, *Statistical Science*, **12(3)**, 133
- Berger, J. & Sellke, T. 1987, *Journal of the American Statistical Association*, **82(397)**, 112
- Bernoulli, J. 1713, *Ars conjectandi, opus posthumum. Accedit Tractatus de seriebus infinitis, et epistola gallicé scripta de ludo pilae reticularis* (Thurneysen Brothers, Basel, Switzerland)
- Bretthorst, G. 2001a, in *American Institute of Physics Conference Series*, Vol. 568, Bayesian Inference and Maximum Entropy Methods in Science and Engineering, ed. A. Mohammad-Djafari, 241–245
- Bretthorst, G. 2001b, in *American Institute of Physics Conference Series*, Vol. 568, Bayesian Inference and Maximum Entropy Methods in Science and Engineering, ed. A. Mohammad-Djafari, 246–251
- Bretthorst, L. 1988, *Bayesian spectrum analysis and parameter estimation* (Springer-Verlag, Berlin, available electronically at [bayes.wustl.edu/glb/book.pdf](http://bayes.wustl.edu/glb/book.pdf))
- Brewer, B. J., Bedding, T. R., Kjeldsen, H., & Stello, D. 2007, *ApJ*, **654**, 551
- Broomhall, A., Chaplin, W. J., Elsworth, Y., Appourchaux, T., & New, R. 2010, *MNRAS*, **406**, 767
- Chaplin, W. J., Appourchaux, T., Elsworth, Y., et al. 2004, *A&A*, **424**, 713
- Chaplin, W. J., Elsworth, Y., Howe, R., et al. 1996, *Sol. Phys.*, **168**, 1
- Christensen-Dalsgaard, J. 1989, *MNRAS*, **239**, 977
- Christensen-Dalsgaard, J. 2004, in *ESA Special Publication*, Vol. 559, SOHO 14 Helio- and Asteroseismology: Towards a Golden Future, ed. D. Danesny, 1

- Christensen-Dalsgaard, J., Carpenter, K. G., Schrijver, C. J., Karovska, M., & the Si Team. 2011, *Journal of Physics Conference Series*, **271**, 012085
- Christensen-Dalsgaard, J. & Gough, D. O. 1982, *MNRAS*, **198**, 141
- Cooley, J. W. & Tuckey, J. W. 1965, *Math. Comp.*, **19**, 297
- Cramer, H. 1946, *Mathematical Methods of Statistics* (Princeton University Press, New Jersey)
- Cuyppers, J. 1987, *Medelingen van de Koninklijke Academie voor Wetenschappen, Letteren en Schone Kunsten van België*, Jg. 49 Nr. 3 (Brussel: Paleis der Academiën), 21
- De Finetti, B. 1937, *Annales de l'Institut Henri Poincaré*, **7**, 1, 1
- De Moivre, A. 1718, *The doctrine of chances of a method of calculating the probability of events in play* (W. Pearson, London, United Kingdom)
- Deheuvels, S., Bruntt, H., Michel, E., et al. 2010, *A&A*, **515**, A87
- Duvall, T. L., J. & Harvey, J. W. 1986, in *Seismology of the Sun and the Distant Stars*, ed. Y. Osaki & H. Shibahashi, *Lecture Notes in Physics*, Berlin Springer Verlag, 105
- Duvall, Jr., T. L., Jefferies, S. M., Harvey, J. W., Osaki, Y., & Pomerantz, M. A. 1993, *ApJ*, **410**, 829
- Fisher, R. A. 1912, *Messenger of Mathematics*, **41**, 155
- Fisher, R. A. 1925, *Proc. Cambridge Philos. Soc.*, **22**, 155
- Fleischmann, M. & Pons, S. 1989, *Journal of Electroanalytical Chemistry and Interfacial Electrochemistry*, **261**, 301
- Fourier, J. 1822, *Théorie analytique de la chaleur* (Firmin-Didot, Paris, France)
- Frieden, B. R. 1983, *Probability, Statistical Optics, and Data Testing: A Problem Solving Approach* (Springer-Verlag, Berlin)
- Fröhlich, C., Andersen, B. N., Appourchaux, T., et al. 1997, *Sol. Phys.*, **170**, 1
- Gabriel, M. 1993, *A&A*, **274**, 931
- Gabriel, M. 1994, *A&A*, **287**, 685
- García, R. A., Régulo, C., Samadi, R., et al. 2009, *A&A*, **506**, 41
- Gaulme, P., Appourchaux, T., & Boumier, P. 2009, *A&A*, **506**, 7
- Gauss, C. F. 1809, *Theoria motus corporum coelestium in sectionibus conicis solem ambientum* (Königliche Gesellschaft der Wissenschaften, Göttingen, Germany)
- Gauss, C. F. 1866, *Nachlass: Theoria interpolationis methodo nova tractata: Werke band 3*, 265-327 (Königliche Gesellschaft der Wissenschaften, Göttingen, Germany)
- Gautschi, A. & Saio, H. 1995, *Ann. Rev. Astron. Astrophys.*, **33**, 75
- Gautschi, A. & Saio, H. 1996, *Ann. Rev. Astron. Astrophys.*, **34**, 551
- Gelly, B., Lazrek, M., Grec, G., et al. 2002, *A&A*, **394**, 285
- Gelman, A. & Meng, X.-L. 1998, *Statistical Science*, **13**(2), 163
- Gibbs, J. 1898, *Nature*, **59**, 200
- Gibbs, J. 1899, *Nature*, **59**, 606
- Gizon, L. & Solanki, S. K. 2003, *ApJ*, **589**, 1009
- Goldstine, H. H. 1977, *A History of Numerical Analysis from the 16th Through the 19th Century* (Springer, New York-Heidelberg-Berlin)
- Gott, J. R. I. 1994, *Nature*, **368**, 106
- Gregory, P. 2005, *Bayesian Logical Data Analysis for the Physical Sciences: A Comparative Approach with 'Mathematica' Support* (Cambridge University Press)
- Gruberbauer, M., Kallinger, T., Weiss, W. W., & Guenther, D. B. 2009, *A&A*, **506**, 1043
- Handberg, R. & Campante, T. L. 2011, *A&A*, **527**, A56
- Harvey, J. 1985, in *Future missions in solar, heliospheric and space plasma physics*, ESA SP-235, ed. E. Rolfe & B. Battick (ESA Publications Division, Noordwijk, The Netherlands), 199-208
- Harvey, J. W., Duvall, Jr., T. L., Jefferies, S. M., & Pomerantz, M. A. 1993, *ASP Conference Series*, **42**, 111
- Hastings, W. K. 1970, *Biometrika*, **57**(1), 97
- Heideman, M. T., Johnson, D. H., & Sidney Burrus, C. 1985, *Archive for History of Exact Sciences*, **34**, 265
- Hill, F., Stark, P., Stebbins, R., et al. 1996, *Science*, **272**, 1292
- Högbom, J. A. 1974, *A&A Sup. Series*, **15**, 417

- Houdek, G., Balmforth, N. J., Christensen-Dalsgaard, J., & Gough, D. O. 1999, *A&A*, **351**, 582
- Jaynes, E. T. 1987, in *Maximum entropy and Bayesian spectral analysis and estimation problems*, ed. C. Smith & G. Erickson (D.Reidel, Dordrecht-The Netherlands), 1
- Jaynes, E. T. 2009, *Probability theory: the logic of Science* (6th edition, Edited by Larry Brethorst, Cambridge University Press, Cambridge, United Kingdom)
- Jeffreys, H. 1939, *Theory of Probability* (Oxford University Press, Oxford, United Kingdom)
- Jeffreys, H. 1946, *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, **186**, 155
- Keshner, M. S. 1982, *IEEE Proceedings*, **70**, 212
- Khinchin, A. Y. 1934, *Mathematische Annalen*, **109**, 604
- Koen, C. 1999, *MNRAS*, **309**, 769
- Komm, R. W., Gu, Y., Hill, F., Stark, P. B., & Fodor, I. K. 1999, *ApJ*, **519**, 407
- Kosovichev, A. G., Schou, J., Scherrer, P. H., et al. 1997, *Sol. Phys.*, **170**, 43
- Kumar, P., Franklin, J., & Goldreich, P. 1988, *ApJ*, **328**, 879
- Laplace, M. 1774, *Mémoires de Mathématique et de Physique*, **T6**, 621
- Legendre, A.-M. 1805, *Nouvelles méthodes pour la détermination des orbites des comètes* (Firmin Didot, Paris, France)
- Libbrecht, K. G. 1992, *ApJ*, **387**, 712
- Mathur, S., Handberg, R., Campante, T. L., et al. 2011, *ArXiv e-prints*
- Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., & Teller, E. 1953, *Journal of Chemical Physics*, **21(6)**, 1087
- Mosser, B., Michel, E., Appourchaux, T., et al. 2009, *A&A*, **506**, 33
- Neyman, J. & Pearson, E. 1933, *Philosophical Transactions of the Royal Society of London. Series A*, **231**, 289
- Nigam, R. & Kosovichev, A. G. 1998, *ApJ Letters*, **505**, L51
- Nigam, R., Kosovichev, A. G., Scherrer, P. H., & Schou, J. 1998, *ApJ Letters*, **495**, L115
- Nyquist, H. 1924, *Bell Syst. Tech. Journal*, **3**, 324
- Parseval des Chênes, M.-A. 1806, *Mémoires présentés à l'Institut des Sciences, Lettres et Arts, par divers savans, et lus dans ses assemblées. Sciences, mathématiques et physiques.*, **1**, 638
- Peligrad, M. & Wu, W. 2010, *The Annals of Probability*, **38**, 2009
- Poincaré, H. 1914, *La Science et l'Hypothèse* (Editions Flammarion, Paris, France)
- Poretti, E., Michel, E., Garrido, R., et al. 2009, *A&A*, **506**, 85
- Press, W. H. & Rybicki, G. B. 1989, *ApJ*, **338**, 277
- Rao, C. 1945, *Bulletin of the Calcutta Mathematical Society*, **37**, 81
- Ritzwoller, M. H. & Lavelle, E. M. 1991, *ApJ*, **369**, 557
- Roberts, D. H., Lehar, J., & Dreher, J. W. 1987, *Astrophysical Journal*, **93**, 968
- Samadi, R. & Goupil, M.-J. 2001, *A&A*, **370**, 136
- Scargle, J. D. 1982, *ApJ*, **263**, 835
- Schou, J. 1992, PhD thesis, Århus Universitet, Denmark
- Schou, J. & Brown, T. M. 1994, *A&A Sup. Series*, **107**, 541
- Schuster, A. 1897, *Proc. Royal Soc. of London*, **61**, 455
- Scott, J. & Berger, J. 2010, *The Annals of Statistics*, **38(5)**, 2587
- Sellke, T., Bayarri, M. J., & Berger, J. 2001, *The American Statistician*, **55**, 62
- Severino, G., Magri, M., Oliviero, M., Straus, T., & Jefferies, S. M. 2001, *ApJ*, **561**, 444
- Shannon, C. E. 1949, *Proceedings of the IRE*, **37**, 10
- Slepian, D. 1978, *Bell System Technical Journal*, **57**, 1371
- Stahn, T. & Gizon, L. 2008, *Sol. Phys.*, **251**, 31
- Sturrock, P. A. & Scargle, J. D. 2009, *ApJ*, **706**, 393
- Tanvir, N. R., Hendry, M. A., Watkins, A., et al. 2005, *MNRAS*, **363**, 749
- Thomson, D. J. 1982, in *Proc. IEEE*, Vol. **70**, 1055–1096
- Toutain, T. & Appourchaux, T. 1994, *A&A*, **289**, 649
- Toutain, T., Appourchaux, T., Baudin, F., et al. 1997, *Sol. Phys.*, **175**, 311
- Toutain, T. & Fröhlich, C. 1992, *A&A*, **257**, 287
- Toutain, T. & Gouttebroze, P. 1993, *A&A*, **268**, 309
- Wehlau, W. & Leung, K. 1964, *ApJ*, **139**, 843

Wiener, N. 1930, Acta Mathematica, **55**, 117

Wilks, S. S. 1938, Annals of Mathematical Statistics, **9**, 60

Yoachim, P., McCommas, L. P., Dalcanton, J. J., & Williams, B. F. 2009, Astrophysical Journal, **137**, 4697